

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
8 July 2004 (08.07.2004)

PCT

(10) International Publication Number
WO 2004/056866 A1

(51) International Patent Classification⁷: **C07K 14/705**,
A61K 38/17, A61P 11/06, C12N 5/12, C12Q 1/68

Strömbergintie 14 B 41, FIN-00380 HELSINKI (FI).
SALMIKANGAS, Paula [FI/FI]; Aurankaari 2 B 8,
FIN-01800 KLAUKKALA (FI).

(21) International Application Number:

PCT/FI2003/000973

(74) Agent: **OY JALO ANT-WUORINEN AB**; Iso
Roobertinkatu 4-6 A, FIN-00120 HELSINKI (FI).

(22) International Filing Date:

19 December 2003 (19.12.2003)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/435,846	20 December 2002 (20.12.2002)	US
60/437,895	3 January 2003 (03.01.2003)	US
60/458,767	26 March 2003 (26.03.2003)	US
60/486,000	9 July 2003 (09.07.2003)	US

(81) Designated States (*national*): AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, EG, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT (utility model), PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): **GENEOS OY** [FI/FI]; Tukholmankatu 2, 2. krs, FI-00250 Helsinki (FI).

(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **LAITINEN, Tarja** [FI/FI]; Myllytie 18, FIN-04400 JÄRVENPÄÄ (FI). **KERE, Juha** [FI/FI]; Lahnatie 1, FIN-02170 ESPOO (FI). **LAITINEN, Lauri, A.** [FI/FI]; Katiskakuja 3 as 6, FIN-02230 ESPOO (FI). **POLVI, Anne** [FI/FI]; Kalliopuontie 21 A 4, FIN-02920 ESPOO (FI). **MÄKELÄ, Siru** [FI/FI]; Sankaritie 7 A 12, FIN-00320 HELSINKI (FI). **VENDELIN, Johanna** [FI/FI]; Kilonrinne 7 B8, FIN-02610 ESPOO (FI). **PULKKINEN, Ville** [FI/FI];

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **ASTHMA SUSCEPTIBILITY LOCUS**

(57) Abstract: The present invention describes a susceptibility locus which is functionally related to asthma. The locus maps within human chromosome 7p15-p14. The invention also describes a novel human gene, GPRA. The invention provides diagnostic methods and materials for analysing allelic variation in said locus and the GPRA gene. The invention also provides polypeptides encoded by GPRA gene and antibodies binding to said polypeptides. The invention further provides pharmaceutical compositions for the treatment of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer.



WO 2004/056866 A1

ASTHMA SUSCEPTIBILITY LOCUS

FIELD OF THE INVENTION

The present invention resides in the field of molecular genetics.

BACKGROUND OF THE INVENTION

We have previously mapped a susceptibility locus for asthma and high total serum Immunoglobulin E level (IgE) to a large region spanning ~19 cM of chromosome 7p15-p14. This locus was the only locus that reached genome wide significance level among the Finnish asthma families (Laitinen, 2001). Linkage was further confirmed among the French-Canadian families recruited based on asthma and in another independent data set from Finland recruited based on high IgE level (Laitinen, 2001). Daniels *et al.* (1996) have reported among Australian and British families six regions of possible linkage including 7p14-p15 and established by simulations that at least some of them are likely to be true positives. Fine mapping of the region revealed bimodal linkage to bronchial hyperreactivity and blood eosinophil count at D7S484 ($P=0.0003$) and D7S669 ($P=0.006$) 63 cM apart (Leaves *et al.* 2002). German (Wjst *et al.* 1999), French (Dizier *et al.* 2000), and Italian families (Malerba *et al.* 2000) have shown some evidence of linkage, but there are also genome scans with inconclusive results (Ober *et al.* 2000; Xu *et al.* 2000; Mathias *et al.* 2001), some of which have been done in ethnic populations other than Caucasian (Yokouchi *et al.* 2000; Xu *et al.* 2001). By comparing genome scans in different immune disorders, it has been suggested that clinically distinct autoimmune diseases may be controlled by a common set of susceptibility genes (Becker *et al.* 1998). 7p15-p14 has also been linked to diseases such as multiple sclerosis (Sawcer *et al.* 1996) and inflammatory bowel disease (Satsangi *et al.* 1996), and genomic regions homologous to human 7p15-p14 have been linked to insulin dependent diabetes (Jacob *et al.* 1992) and inflammatory arthritis (Remmers *et al.* 1996) in rat models.

SUMMARY OF THE INVENTION

The present invention provides an isolated, purified asthma locus-1 (AST1) nucleic acid and a complement or a fragment thereof. The invention also provides nucleic acids comprising at least one single nucleotide polymorphism and/or deletion/insertion polymorphism in different positions in asthma locus-1. One object of the invention is to provide vectors, host cells, primers, and probes comprising asthma locus-1 nucleic acid according to the invention. The present invention is also related to a method for the diagnosis of a single nucleotide polymorphism or a deletion/insertion polymorphism in asthma locus-1 according to the

invention in a human, which method comprises determining the sequence of the nucleic acid of the human at one or more of positions in asthma locus-1 and determining the status of the human by reference to polymorphism in asthma locus-1. The invention is also related to a kit for use in the diagnostics of asthma and other IgE mediated allergic diseases or in assessing the predisposition of an individual to asthma and other IgE mediated allergic diseases. The invention is further related to a method for identifying a mutation, which increases individual's susceptibility to develop asthma and other IgE mediated allergic diseases. The invention also provides a transgenic animal comprising asthma locus-1 nucleic acid according to the invention.

The invention provides an isolated GPRA polypeptide comprising an amino acid sequence that has at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over the entire length of the selected SEQ ID No: when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix.

The invention further provides an isolated GPRA polypeptide comprising at least 10 contiguous amino acids from amino acids 343-377 of B-long (SEQ ID NO:5).

The invention further provides an isolated GPRA polypeptide comprising an amino acid sequence that has at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over a sequence comparison window of at least 40 amino acids when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix provided that the polypeptide includes a variant amino acid encoded by a variant form shown in Table 7.

The invention further provides an isolated GPRA polypeptide comprising the amino acid sequence of SEQ ID NO:5 provided that the sequence contain as an amino acid substitution of Asn instead of Ile at codon position 107, Arg instead of Ser at codon position 241, and/or Thr instead of Ile at codon position 366.

The invention further provides an an isolated nucleic acid encoding a GPRA polypeptide as defined above.

The invention further provides an isolated nucleic acid that hybridizes under highly stringent conditions to any of SEQ ID NOS: 1, 4, 6, 8, 10, 12, and 14 without hybridizing under the same highly stringent conditions to SEQ ID NO:2, wherein the highly stringent conditions are $6 \times$ NaCl/sodium citrate (SSC) at about 45 °C for a hybridization step, followed by a wash of $2 \times$ SSC at 50 °C.

The invention further provides an isolated nucleic acid having a sequence that is at least 90 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 4, 6, 8, 10, 12, and 14 over the entire length of the selected SEQ ID NO when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N= -4.

The invention further provides an isolated nucleic acid having a sequence that is at least 80 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 1, 2, 4, 6, 8, 10, 12, and 14 over a sequence comparison window of at least 100 nucleotides when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N= -4 provided that the nucleic acid includes a polymorphic site occupied by a variant form as shown in Table 3 or Table 7.

The invention further provides an isolated nucleic acid having a sequence that is at least 80 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 1, 2, 4, 6, 8, 10, 12, and 14 over a sequence comparison window of at least 100 nucleotides when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N= -4 provided that the nucleic acid includes a polymorphic site occupied by a reference form designated with a * in Table 7.

The invention further provides an isolated genomic DNA molecule or a minigene having at least one intronic sequence and encoding a GPRA polypeptide that has at least 80 % sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over a region at least 40 amino acids in length when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix .

The invention further provides an antibody that specifically binds to an epitope within amino acids 343-377 of B-long (SEQ ID NO:5) or amino acids 332-366 of B-short (SEQ ID NO:7).

The invention further provides a method of preventing or treating asthma, other IgE mediated disease or cancer. The method comprises administering to a patient suffering from or at risk of asthma, other IgE mediated disease or cancer an effective amount of a modulator of a GPRA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS 3, 5, 7, 9, 11, 13.

The invention further provides a method of identifying a modulator of a GPRA polypeptide. The method comprises contacting a cell expressing a GPRA polypeptide with an agent; and determining whether the agent modulates expression of the GPRA polypeptide and/or signal transduction through the GPRA polypeptide, wherein the GPRA polypeptide is defined as above.

The invention further provides an method of determining risk of asthma, other IgE mediated disease or cancer. The method comprises determining whether the individual has a variant polymorphic form in a GPRA gene, wherein presence of the variant polymorphic form indicates risk of asthma, other IgE mediated disease or cancer.

The invention further provides a method for identifying a polymorphic site correlated with a disease selected from the group consisting of asthma, other IgE-mediated disease and cancer or susceptibility thereto. The method comprises identifying a polymorphic site within a GPRA gene; and determining whether a variant polymorphic form occupying the site is associated with the disease or susceptibility thereto.

The invention further provides a primer or probe nucleic acid that hybridizes under highly stringent conditions to a segment of SEQ ID NO:1, 2 or 4 or a variant form thereof differing from SEQ ID NO: 1, 2 or 4 at a position shown in Table 3 or Table 7, wherein the segment includes or is immediately adjacent to a polymorphic site shown in Table 3 or Table 7.

The invention further provides a transgenic animal comprising a GPRA and/or AAA1 nucleic acid.

The invention further provides a transgenic animal disposed to develop a characteristic of asthma, other IgE-mediated disease or cancer in which an endogenous a GPRA gene encoding a cognate form of a GPRA polypeptide defined by any of SEQ ID NOS: 3, 5, 7, 9, 11, 13 and 15 is functionally disrupted to prevent expression of a gene product.

The invention further provides a kit for use in diagnosing or assessing predisposition to asthma, other IgE-mediated disease or cancer. The kit comprises a container; and in the container, a compound, preferably labeled, capable of detecting a polymorphic form at a polymorphic site in a susceptibility locus for asthma as defined by SEQ ID NO:1, 2 or 4.

The invention further provides an isolated AAA1 polypeptide comprising an amino acid sequence that has at least at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 over the entire length of the selected SEQ ID No: when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix.

The invention further provides an isolated nucleic acid encoding the AAA1 polypeptide as defined above.

The invention further provides an isolated nucleic acid having a sequence that is at least 80% identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 over the entire length of the selected SEQ ID NO when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N= -4.

The invention further provides an isolated nucleic acid having at least 20 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40.

The invention further provides an isolated genomic DNA molecule or a minigene having at least one intronic sequence and encoding an AAA1 polypeptide that has at least 80 % sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 over the entire length of the selected SEQ ID NO. when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix .

The invention further provides an antibody that specifically binds to a polypeptide selected from the group consisting of SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

The invention further provides a method of preventing or treating asthma, other IgE-mediated disease or cancer. The method comprises administering to a patient suffering from or at risk of asthma, other IgE-mediated disease or cancer an effective amount of a modulator of an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

The invention further provides a method of identifying a modulator of an AAA1 polypeptide. The method comprises contacting an AAA1 polypeptide with an agent; and determining whether the agent binds to the AAA1 polypeptide, modulates expression of the AAA1 polypeptide or modulates activity of the AAA1 polypeptide, wherein the AAA1 polypeptide comprises an amino acid sequenced as defined by any of SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

The invention further provides a method of determining risk of asthma, other IgE mediated disease or cancer, comprising determining whether the individual has a variant polymorphic form in an AAA1 gene, wherein presence of the variant polymorphic form indicates risk of asthma, other IgE mediated disease or cancer.

The invention further provides a method for identifying a polymorphic site correlated with a disease selected from the group consisting of asthma, other IgE mediated disease or cancer or susceptibility thereto. The method comprises identifying a polymorphic site within an AAA1 gene, and determining whether a variant polymorphic form occupying the site is associated with the disease or susceptibility thereto.

The invention provides a primer or probe nucleic acid of nucleotides that hybridizes under highly stringent conditions to a segment of SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 or a variant form thereof differing from SEQ ID NOS.16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 at a single polymorphic site.

The invention further provides a transgenic animal disposed to develop a characteristic of asthma, other IgE-mediated disease or cancer in which an endogenous a AAA1 gene encoding a cognate form of an AAA1 polypeptide defined by any of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 is functionally disrupted to prevent expression of a gene product.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1: Physical map across the linkage region showing the organization of exons of known genes and microsatellite markers genotyped (above), and the organization of genomic BAC clones and contigs (below).

Fig. 2: Physical map across the critical (SNP509783-SNP638799, a total of 129,017 bp) and the flanking regions showing the organization of all markers genotyped in phase three (above), and the organization of genomic BAC clones (below).

Fig. 3: Genomic localization of the susceptibility haplotype (AST1, gray region) of 129 kb for asthma related traits in relation to verified exons of *GPRA* (G protein coupled receptor for asthma susceptibility). The cDNA probe used in the Northern blotting is shown in shaded gray. Locations of PCR primers used in the identification of the gene structures and nested PCR primers (in bold) used in cloning of the full length cDNA of splice variants A and B are shown. SNPs in the exons of *GPRA* and their potential amino acid changes are shown below. Both ends of the susceptibility haplotype are shown with arrows as physical positions in the genomic contig NT_000380.

Figs. 4A, 4B1, 4B2, 4C, 4D, 4E, and 4F: Deduced cDNAs (SEQ ID NOS: 2, 4, 6, 8, 10, 12, and 14) and amino acid sequences (SEQ ID NOS: 3, 5, 7, 9, 11, 13, and 15) of the alternative splice variants (A-F) of *GPRA*. Exon borders are indicated by vertical lines. The initiation and stop codons are shown in bold.

Figs. 5A, 5B, 5C, 5D, 5E and 5F: Genomic structures of the alternative splice variants (A-F) of *GPRA*. The exons (gray boxes) are depicted by numbered boxes (E1-E9), the length of the exons are indicated below, and the length of the separating introns above. The translation-initiation site and the stop codons are indicated by arrows. The white boxes indicate the 5' and 3' UTR regions and the shaded box in the B variant exon 3 shows the alternative slicing of the long and short form.

Figs. 6A, 6B, 6C, 6D, 6E and 6F: Predicted 7TM structures of the alternative splice variants of *GPRA* (SEQ ID NOS: 3, 5, 7, 9, 11, 13, and 15). For the splice variants A and B_{long}, cytoplasmic loops (=cytoloop) and extracellular loops (=exoloop) are indicated above the

amino acid chain, transmembrane regions are underlined and the conserved amino acids, characteristic for the GPR family A, are bolded. For the other splice variants (B_{short} -F) the number of predicted transmembrane regions varied from one to six.

Fig. 7: Predicted topology of the seven transmembrane structure of the protein encoded by the *GPRA* A splice variant by TMpred.

Fig. 8: Expression of *GPRA* variants in the human lung epithelial carcinoma cell line NCI-H358. RT-PCR was performed by using variant A, B, and C specific primers. PCR-products were separated on ethidium bromide stained 1.5 % agarose gel.

Fig. 9: Expression of *GPRA* by Northern blot analysis with a 470 bp cDNA probe (comprising the exons E2a and E2b) in human placental tissue. Four transcripts were detected that were approximately 6.0 kb, 4.5 kb, 1.9 kb, and 1.0 kb in size. The expression of β -actin is shown as control.

Figs. 10A, 10B, 10C, 10D, 10E, and 10F: Immunohistochemical analysis of *GPRA* A in human bronchial (10A, 10B), lung (10C), and colon (10D, 10E), and skin (10D) tissues. Paraffin sections were stained using immunoperoxidase technique with *GPRA* variant A specific antibody (10A, 10C - 10D) or pre-immune sera (10B). Strong immunostaining of *GPRA* A in bronchus (10A) and colon (10E) was recorded in smooth muscle cell layer in bronchial walls (thick arrows) and in arterial walls (thin arrows) and subepithelially (arrowheads) in colon, respectively. In alveolar walls and alveolar macrophages (asterisk) intense staining was detected (10C). In colon (10D) and skin (10F) tissues, basal cell layer of the epithelium (white arrows) stained positive for *GPRA* A. Staining of the corresponding sections with the pre-immune serum did not show specific reactivity (10B). Original magnification x400.

Figs. 11A, 11B, 11C, 11D, 11E, and 11F: Immunohistochemical analysis of *GPRA* B in human bronchial (11A, 11B), lung, small intestine (11D), colon (11E), and skin (11F) tissues. Paraffin sections were stained with the immunoperoxidase technique with *GPRA* variant B specific antibody (11A, 11C-D) or preimmune serum (11B). Strong immunostaining of *GPRA* B was recorded in the epithelium of the bronchial wall (11A), small intestine (11D), colon (11E), and skin (11F) tissues (arrows). In lung intense staining of alveolar wall and alveolar macrophages (asterisk) was detected (11C). Staining of the corresponding sections with the preimmune serum did not show specific reactivity (11B). Original magnification x 400.

Figs. 12A and 12B: In Western blot analysis of several human tissues, the A variant specific antibody recognized four major polypeptide bands at approximately 40, 42, 44, and 50 kDa (12A). 50 kDa band was detectable in skeletal muscle whereas uterine muscle, colon epithelium and prostate showed similar expression patterns of *GPRA A* at approximately 50, 44, 40 kDa. The most intensive *GPRA A* expression was recorded in colon muscle at 42 kDa. The B variant specific antibody recognized two major polypeptide bands at approximately 25 and 39 kDa with the most intensive *GPRA* expression in kidney (12B).

Fig. 13: Genomic structure of AAA1 showing both the exon (size below) and intron (size above) structures of the gene. AAA1 (asthma associated alternatively spliced gene 1) shows complex splicing, two alternative starting methionines and six alternative stop codons for the predicted polypeptide was identified. Location of the primers used in cloning of different splice variants are shown below. Gray area shows the location of AST1.

Fig. 14: I-XII different splice variants of AAA1, I-XI are full length cDNAs.

Fig. 15: Sequence alignments of the predicted polypeptides encoded by the I-XI splice variants of AAA1 (CLUSTAL W program). The conserved region of the AAA1 encoded polypeptides is in bold.

Fig. 16: Human multiple tissue expression array (BD, Clontech) hybridized with the probe specific for AAA1 (a mixture of multiple splice variants) (upper panel) and location of studied RNAs in the dot blot (lower panel).

Fig. 17: Human fetal multiple tissue northern blot (BD, Clontech) hybridized with the probe specific for AAA1 (a mixture of multiple splice variants) identifies several alternative transcripts (arrows).

Fig. 18: Tissue specific expression AAA1. RT-PCR with the variants I, IV, VI, X, and XI specific primer pairs in liver (Li), lung (Lu), testis (Te), and kidney (Ki) shows different patterns of transcripts in different tissues.

Fig. 19: Variable alternative splicing for AAA1 depending on genotype. RT-PCR spanning exons 6 to 10b of AAA1 was performed on lymphoblastic RNA samples genotyped for AST1. Only the non-carrier of AST1 processes normal amount of the exon 6-10b transcript, whereas a homozygote and heterozygotes show either absent transcript or smaller splice variants. Beta-actin was used as control in parallel amplifications.

Fig. 20: Nucleic (SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40) and amino acid sequences (SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41) of splice variants of AAA1.

Figs. 21A and 21B: 21A. *In vitro* translation of AAA1 gene using rabbit reticulocyte lysate translation machinery with S³⁵-labelled methionine in the reaction mixture. 21B. Specificity of the AAA1 antibody (anti-AAA1). Glutathione S-transferase (GST) -fusion proteins for AAA1 were expressed in *E. coli* as GST fusion proteins (GST-AAA1).

DEFINITIONS

Unless defined otherwise, all technical and scientific terms have the same meaning as is commonly understood by one of skill in the art to which this invention is related. The definitions below are presented for clarity.

"Isolated" when referred to a molecule, refers to a molecule that has been identified and separated and/or recovered from a component of its natural environment and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide. The term "isolated" does not refer to genomic or cDNA libraries, whole cell total or mRNA preparations, genomic DNA preparations (including those separated by electrophoresis and transferred onto blots), sheared whole cell genomic DNA preparations or other compositions where the art demonstrates no distinguishing features of the polynucleotide sequences of the present invention.

"Nucleic acid", includes DNA molecules (*e.g.* cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs. The nucleic acid or nucleic acid molecule may be single-stranded or double-stranded, but preferably comprises double-stranded DNA. Preferred nucleic acids of the invention include segments of DNA, or their complements including any one of the polymorphic sites shown in Table 3, 7 or 12. The segments are usually between 5 and 100 contiguous bases, and often range from 5, 10, 12, 15, 20, or 25 nucleotides to 10, 15, 30, 25, 20, 50 or 100 nucleotides. Nucleic acids between 5-10, 5-20, 10-20, 12-30, 15-30, 10-50, 20-50 or 20-100 bases are common. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in Table 3, 7 or 12. For brevity in Table 3, the symbol T is used to represent both thymidine in DNA and uracil in RNA. Thus, in RNA oligonucleotides, the symbol T should be construed to indicate a uracil residue. Unless otherwise apparent from the context,

reference to a SEQ ID NO. of the invention, refers to the strand shown in the sequence listing for that SEQ ID NO., the perfect complementary strand thereof or a duplex of the two. Thus, for example, reference to primers hybridizing to SEQ ID NO:1 includes primers hybridizing to the strand shown in the sequence listing and the complementary strand.

"Isolated nucleic acid" is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an isolated nucleic acid is free of sequences that naturally flank the nucleic acid (*i.e.* sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, isolated asthma locus-1 molecules can contain less than about 5 kb, 4 kb, 3kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an isolated nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

The term "fragment" or "segment" as applied to a longer nucleic acid, may ordinarily be at least about 10 contiguous nucleotides of the longer nucleic acid in length, typically, at least about 20 contiguous nucleotides, more typically, from about 20 to about 50 contiguous nucleotides, preferably at least about 50 to about 100 contiguous nucleotides, even more preferably at least about 100 contiguous nucleotides to about 300 contiguous nucleotides, yet even more preferably at least about 300 to about 400, and most preferably, the nucleic acid fragment will be greater than about 500 contiguous nucleotides in length.

"Oligonucleotide" comprises a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction or other application. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid.

"Variant" refers to a polynucleotide differing from the polynucleotide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely

similar, and, in many regions, identical to the polynucleotide of the present invention. Allelic variants of a gene refer to variant forms of the same gene between different individuals of the same species. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms. Cognate forms of a gene refers to variation between structurally and functionally related genes between species. For example, the human gene showing the greatest sequence identity and closest functional relationship to a mouse gene is the human cognate form of the mouse gene. Thus, for example, the invention includes primate, bovine, ovine, murine, and avian cognate forms of the human GPRA and AAA1 genes.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (*e.g.*, sequences that vary in less than 1/100 or 1/1000 members of the populations). A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

A set of polymorphisms means at least 2, and sometimes 5, or more of the polymorphisms shown in Table 3, 7 or 12.

"Stringency". Highly stringency conditions are well-known in the art, *e.g.* 6×NaCl/sodium citrate (SSC) at about 45 °C is applied for a hybridization step, followed by a wash of 2×SSC at 50 °C or , *e.g.*, alternatively hybridization at 42 °C in 5×SSC, 20 mM NaPO₄, pH 6.8, 50% formamide; and washing at 42 °C in 0.2×SSC. These conditions can be varied empirically based on the length and the GC nucleotide base content of the sequences to be hybridized, or based on formulas for determining such variation (See, for example, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", Second Edition, pages 9.47-9.51, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (1989)). A low stringency is

defined herein as being in 4-6 X SSC/0.1-0.5% w/v SDS at 37-45 degree of C for 2-3 hours. Depending on the source and concentration of nucleic acid involved in the hybridization, alternative conditions of stringency may be employed such as medium stringent conditions which are considered herein to be 1-4 X SSC/0.25-0.5% w/v SDS at 45 degree of Celsius for 2-3 hours or highly stringent conditions considered herein to be 0.1-1 X SSC/0.1% w/v SDS at 60 degree of Celsius for 1-3 hours.

Hybridization probes are capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include nucleic acids, peptide nucleic acids, as described in Nielsen *et al.*, Science 254, 1497-1500 (1991). "Probes" are nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or many (*e.g.*, 6000 nt) depending on the specific use. Often probes have 15-50 nucleotides. Probes are used to detect identical, similar, or complementary nucleic acid sequences. Longer length probes can be obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies. Probes also hybridize to nucleic acid molecules in biological samples, thereby enabling immediate applications in chromosome mapping, linkage analysis, tissue identification and/or typing, and a variety of forensic and diagnostic methods of the invention.

The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30, 40 or 50 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

"Vector" means any plasmid or virus encoding an exogenous nucleic acid. The term should also be construed to include non-plasmid and non-viral compounds which facilitate transfer of nucleic acid into virions or cells, such as, for example, polylysine compounds and the like. The vector may be a viral vector which is suitable as a delivery vehicle for delivery of the nucleic acid encoding the desired protein, or mutant thereof, to a cell, or the vector may be a non-viral vector which is suitable for the same purpose. Examples of viral and non-viral vectors for delivery of DNA to cells and tissues are well known in the art and are described, for example, in Ma *et al.* (1997, Proc. Natl. Acad. Sci. U.S.A. 94:12744-12746). Examples of viral vectors include, but are not limited to, a recombinant vaccinia virus, a recombinant adenovirus, a recombinant retrovirus, a recombinant adeno-associated virus, a recombinant avian pox virus, and the like (Cranage *et al.*, 1986, EMBO J. 5:3057-3063; International Patent Application No. W094/17810, published August 18, 1994; International Patent Application No. W094/23744, published October 27, 1994). Examples of non-viral vectors include, but are not limited to, liposomes, polyamine derivatives of DNA, and the like.

"Recombinant polynucleotide" refers to a polynucleotide having sequences that are not naturally joined together. An amplified or assembled recombinant polynucleotide may be included in a suitable vector, and the vector can be used to transform a suitable host cell.

"Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, horses, cats, or cows. Preferably, the mammal is human.

Linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers that are physically-linked on the same chromosome. Loci occurring within 50 centimorgan of each other are linked. Some linked markers occur within the same gene or gene cluster.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency

of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair.

Linkage disequilibrium (LD) or allelic association means the preferential association of a particular allele or genetic marker with a specific allele, or genetic marker at a nearby chromosomal location more frequently than expected by chance for any particular allele frequency in the population. For example, if locus X has alleles a and b, which occur equally frequently, and linked locus Y has alleles c and d, which occur equally frequently, one would expect the haplotype ac to occur with a frequency of 0.25 in a population of individuals. If ac occurs more frequently, then alleles a and c are considered in linkage disequilibrium. Linkage disequilibrium may result from natural selection of certain combination of alleles or because an allele has been introduced into a population too recently to have reached equilibrium (random association) with between linked alleles.

A marker in linkage disequilibrium with disease predisposing variants can be particularly useful in detecting susceptibility to disease (or association with other sub-clinical phenotypes) notwithstanding that the marker does not cause the disease. For example, a marker (X) that is not itself a causative element of a disease, but which is in linkage disequilibrium with a gene (including regulatory sequences) (Y) that is a causative element of a phenotype, can be used to indicate susceptibility to the disease in circumstances in which the gene Y may not have been identified or may not be readily detectable. Younger alleles (*i.e.*, those arising from mutation relatively late in evolution) are expected to have a larger genomic segment in linkage disequilibrium. The age of an allele can be determined from whether the allele is shared between different human ethnic human groups and/or between humans and related species.

Unless otherwise apparent from the context, any embodiment, element or feature of the invention can be used in combination with any other.

“Analog” are nucleic acid sequences that have a structure similar to, but not identical to, the native compound but differ from it in respect to certain components. Usually, an analog has the same or similar function to the native compound. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids of the invention include, nucleic acids that are substantially identical to an exemplified SEQ ID NO. and/or capable of hybridizing to the complement of the sequence under highly stringent, moderately stringent, or low stringent conditions.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptides, refer to two or more sequences or subsequences that are the same or have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm such as those described below for example, or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least 80, preferably at least 85%, more preferably at least 90%, 95%, 99% or higher nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm such as those described below for example, or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least 40 residues (*i.e.*, amino acids or nucleotides) in length, preferably over a longer region than 50 residues, more preferably at least about 90-100 residues, and most preferably the sequences are substantially identical over the full length of the sequences being compared, such as the coding region of a nucleotide for example. For example, when a SEQ ID NO. of the invention serves as a reference for comparison with an object nucleic acid, the comparison is preferably performed over the entire length of the SEQ ID NO. of the invention.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel *et al.*, *supra*).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. For identifying whether a nucleic acid or polypeptide is within the scope of the invention, the default parameters of the BLAST programs are suitable. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. The TBLASTN program (using protein sequence for nucleotide sequence) uses as defaults a word length (W) of 3, an expectation (E) of 10, and a BLOSUM 62 scoring matrix. (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, *e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under highly stringent conditions. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence. The phrase "hybridizing specifically to", refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under highly stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA.

The phrases "specifically binds" refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated conditions, a specified ligand binds preferentially to a particular protein and does not bind in a significant amount to other proteins present in the sample. A molecule such as antibody that specifically binds to a protein often has an association constant of at least 10^6 M^{-1} or 10^7 M^{-1} , preferably 10^8 M^{-1} to 10^9 M^{-1} , and more preferably, about 10^{10} M^{-1} to 10^{11} M^{-1} or higher. An antibody that specifically binds to one segment of a protein (*e.g.*, residues 1-10) does not bind to other segments of the protein not included within or overlapping the designated segment.

The term AAA1 gene means nucleotides 163615–684776 of the human genomic contig NT_000380 and allelic or species variants thereof.

The term GPRA gene means nucleotides 471478-691525 of the human genomic contig NT_000380 and allelic or species variants thereof.

A "pharmacological" activity means that an agent exhibits an activity in a screening system that indicates that the agent is or may be useful in the prophylaxis or treatment of a disease. The screening system can be in vitro, cellular, animal or human. Agents can be described as having pharmacological activity notwithstanding that further testing may be required to establish actual prophylactic or therapeutic utility in treatment of a disease.

Pulmonary diseases associated with lower airways obstruction are classified as chronic obstructive pulmonary disease (COPD) or asthma on the basis of the clinical picture. "Asthma" is a chronic inflammatory disease of the airways causing variable airflow obstruction that is often reversible either spontaneously or with treatment. Asthma is highly correlated with other allergic, IgE mediated diseases, such as allergic rhinitis and dermatitis. COPD is characterized by slowly progressing, mainly irreversible airway obstruction and decreased expiratory flow rate.

DETAILED DESCRIPTION OF THE INVENTION

I. General

This invention is based in part on the discovery and characterization of a novel susceptibility locus for asthma and other IgE mediated diseases, two overlapping genes within the locus, termed GPRA and AAA1, and proteins encoded by the genes. The genes have opposite transcriptional orientations and share intronic regions although not exons. The locus maps within human chromosome 7p15-p14.

Fig. 3 shows the structure of the human GPRA gene. The gene has nine exons, two of which have alternate forms, and eight introns. The total length of the gene from the first base of the first exon to the last base of the last exon is 220047 bp and the coordinates are 471478 (beginning) 691525 (end) in NT_000380. Part of the genomic sequence of GPRA including part of intron 2, exon 3, intron 3, exon 4 and part of intron 4 defines a sublocus referred to as AST1. This locus is 129,017 bp in length. The locus refers to the part of human chromosome 7 having the 129,017 bp of SEQ ID NO:1 or allelic or species variants thereof.

Fig. 13 shows the structure of the human AAA1 gene. The gene has nineteen exons, nine of which have alternate forms, and 18 introns. The total length of the gene from the first base of the first exon to the last base of the last exon is 521,161 bp and the coordinates are 163,615 (beginning) 684776 (end) in NT_000380. Part of the genomic sequence of AAA1 including part of intron 2, exon 3 (a and b), intron 3, exon 4 (a and b), intron 4, exon 5 (a and b), intron 5, exon 6, intron 6, exon 7 (a and b), intron 7, exon 8, intron 8, exon 9, intron 9, exon 10 (a and b) and part of intron 10 occurs within the AST1 locus.

The present application shows that the GPRA and AAA1 genes and AST-1 locus within them are genetically linked and associated with IgE diseases, such as asthma. In addition, the invention provides a collection of polymorphic sites including SNPs (single nucleotide polymorphisms) and DIPs (deletion/insertion polymorphisms) within the locus. Tables 3, 7 and 12 show variant and reference (or wildtype) forms occupying these sites. The variant forms are associated with asthma, other IgE mediated diseases, and/or immune-mediated diseases including autoimmune diseases, and/or cancer and are useful in detection, diagnosis, treatment and prophylaxis thereof.

The AST-1 locus was localized by hierarchical LD mapping in a systemic search for a shared haplotype among the individuals with high IgE serum level across a 19 Mb region that was first identified by a genome wide screen for asthma genes (Laitinen *et al.* 2001). This analysis defined the AST-1 locus between the markers SNP509783 and SNP638799 and provided statistically significant evidence that specific haplotype of 129kb within the linkage peak is a risk factor for asthma related traits (SEQ ID NO:1). The best haplotype patterns of polymorphic sites in GPRA (Table 8) were found with the frequency of 13-18% among the high IgE associated chromosomes compared to that of 3-7% among control chromosomes suggesting the risk ratio of 3.9 [95%CI: 1.8-8.6] for high IgE level (38/304 among affected vs. 7/220 control chromosomes). The best observe associations for polymorphic sites in AAA1 reached the χ^2 values of 8.9-13.6 All markers identified between the positions 509,783-638,799 in the genomic contig NT_000380 were found to be in strong LD and therefore useful for diagnostic purposes.

II. GPRA Polypeptides

The human GPRA gene can be expressed as at least seven different splice variants termed A, B-long, B-short, C, D, E, and F. The cDNAs encoding each form and amino acid sequence of each form are shown in Fig. 4 (SEQ ID NOS:2-15). The different splicing events that lead to the different forms are shown in Fig. 5. It can be seen that the B-long and B-short forms differ from the A form in that the B forms have a different C-terminus encoded by exon 9b, whereas the C terminus of the A form is encoded by exon 9A. The B-short form (SEQ ID NOS:6 and 7) differs from the B-long form in that the B-short form has a deletion of 10 amino acids encoded by exon III. The C-form (SEQ ID NOS:8 and 9) contains a segment encoded by exon 2b which is not found in other forms. The C-form lacks any segments encoded by exons 3-9. The D form (SEQ ID NOS:10 and 11) lacks a segment encoded by exon 3, causing a frameshift, whereby exons 4 and 5 are read in a different reading frame causing truncation within exon 5. Form E (SEQ ID NOS:12 and 13) lacks exon 4 causing a frameshift as a result of which exon 5 is read in a different reading frame causing truncation within exon 5. Form F (SEQ ID NOS:14 and 15) lacks exons 3 and 4.

Fig. 6 shows the expected structural characteristics of the different splice variants. The A form includes 371 amino acids and has an extracellular N-terminal domain, seven transmembrane domains and an intracellular C-terminal domains (SEQ ID NO:3). The transmembrane domains are separated by alternating cytoplasmic and extracellular loops. The B-long form (377 amino acids, SEQ ID NO:5) has a similar structure. The predicted size of the extracellular N-terminal domain is about 50 amino acids anticipating suggesting that the endogenous ligand that interacts with GPRA is a small molecule or peptide. The B-short, C, D, E and F variants of GPRA (SEQ ID NOS: 5, 7, 9, 11, 13, and 15, respectively) lack the 7TM structure present in variants A and B-Long.

Fig. 7 shows the arrangement of the A or B-Long form in a cell membrane. GPRA falls within the conserved class A of G-protein coupled receptors (GPRs) (Rana *et al.* 2001, Johnson *et al.* 2000). GPRs are a large and functionally diverse protein superfamily that form a seven-transmembrane (7TM) helices bundle with alternating extracellular and intracellular loops (see WO 01/18206). Class A contains most well known members of the GPRs such as vasopressin, oxytocin, and bovine rodopsin receptors. For the latter the crystallographic structure is available (Palczewski *et al.* 2000). GPRA is expressed in human tissues consistent with its role in asthma and IgE mediated disease. Transcripts of *GPRA* variants A, B, and C were expressed in several tissues including lung and NCI-H358 cell line that

represent cells of broncho-epithelial origin (Fig. 8 and 9). Based on immunohistochemistry, the A variant of the GPRA protein was expressed in smooth muscle cells of bronchi and arterial wall in human lung as subepithelial of smooth muscle cells in colon (Fig. 10) and the B (-long) variant in the epithelium of bronchi and colon (Fig. 11).

A class of GPRA polypeptides can be defined with reference to the exemplary polypeptides defined by SEQ ID NOS: 3, 5, 7, 9, 11, 13 and 15. The class includes allelic, cognate and induced variants of the exemplified polypeptides. Preferred polypeptides shows substantial sequence identity to one or more of the exemplified polypeptides. The class also includes fragments of the exemplified polypeptides having at least 6, 10, 20, 50 100, 200, or 300 contiguous amino acids from one of the above SEQ ID NOS. Some fragments include one or more isolated domains from one of the exemplified polypeptides, *e.g.*, any of the seven transmembrane domains, any of three extracellular loops, any of three intracellular loops, an N-terminal domain and a C-terminal domain. Some polypeptides comprise the amino acid of one of the above SEQ ID NOS. provided that up to 1, 2, 5, 10, 20, 30 or 34 amino acids can be inserted, deleted or substituted relative to the SEQ ID NO. Some GPRA polypeptides comprise an epitope found in one of the exemplified SEQ ID NOS. but not in others. For example, B-Long (SEQ ID NO:5) has a C-terminal epitope of 35 amino acids not found in any of the other forms. Some GPRA polypeptides contain a polymorphic site shown in Table 7 occupied by a variant form.

Preferred GPRA polypeptides are receptors that can be activated to transduce a signal by the same ligand as the exemplified GPRA polypeptide of SEQ ID NOS: 3, 5, 7, 9, 11, 13, and 15. Binding of a ligand or analog to a G-protein receptor induces an alteration in receptor G-protein interaction. The receptor G-protein interaction releases GDP specifically bound to the G protein and permits the binding of GTP, which activates the G protein. Activated G-protein dissociates from the receptor and activates an effector protein, which in turn regulates intracellular levels of second messengers, such as adenylyl cyclase, guanylyl cyclase, and phospholipase C. Signal transduction can be monitored in cells transfected with a reporter construct consisting of several copies of the consensus cAMP response element (Arias *et al.*, *Nature* 370, 226-229 (1994)), a minimal tk promoter, and a secreted alkaline phosphatase gene (Clontech) or can be monitored in changes of the concentrations of intracellular calcium.

III. AAA1 Polypeptides

The human AAA1 gene can be expressed as at least thirteen different splice variants termed IA, IB, II, III, IVA, IVB, V, VI, VII, VIII, IX, X, and XI. The cDNAs encoding each form and amino acid sequence of each form are shown in Fig. 20 (SEQ ID NOS:16-41). The different splicing events that lead to the different forms are shown in Fig. 14. It can be seen that all splice variants have exon VI and the same core peptide sequence (AYVRRNAGRQFSHCNLHAHQFLVRRKQ). The AAA1 gene is expressed predominantly in the testis, brain, placenta, lung, heart, skeletal muscle, kidney, liver, fetal liver and fetal lung. In addition to the variants described in Fig. 14, using 3'-RACE we have found also several other variants expressed in different tissues. They all share the exon six.

A class of AAA1 polypeptides can be defined with reference to the exemplary polypeptides defined by SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41. The class includes allelic, cognate and induced variants of the exemplified polypeptides. Preferred polypeptides shows substantial sequence identity to one or more of the exemplified polypeptides. The class also includes fragments of the exemplified polypeptides having at least 6, 10, 20, or 50 contiguous amino acids from one of the above SEQ ID NOS. Some fragments include or consist of the above core peptide sequence. Some polypeptides comprise the amino acid of one of the above SEQ ID NOS. provided that up to 1, 2, 5, 10, or 20 amino acids can be inserted, deleted or substituted relative to the SEQ ID NO. Some AAA1 polypeptides comprise an epitope found in one of the exemplified SEQ ID NOS. but not in others. For example, only the variant 1B contains the peptide sequence encoded by exon III.

IV. GPRA Nucleic Acids

The present invention provides isolated nucleic acids encoding GPRA polypeptides. The nucleic acids can be, for example, genomic, cDNA, RNA or mini-gene (*i.e.*, a hybrid of genomic and cDNA). Genomic and mini-gene nucleic acids contain at least one intronic segment from a GPRA gene, meaning the section of human chromosome 7p15-p14 having the structure shown in Fig. 3, or allelic and cognate variants thereof. Nucleic acids of the invention can include coding regions, intronic regions, 3' untranslated regions, 5' untranslated regions, 3' flanking regions, 5' flanking regions, enhancers, promoters and other regulatory sequences. A class of GPRA nucleic acids can be defined with reference to exemplified GPRA polypeptide and nucleic acid sequences. The class includes allelic, cognate and

induced variants of the exemplified sequences, as well as nucleotide substitutions, which due to the degeneracy of the genetic code, do not effect the amino acid sequence of an encoded polypeptide. The class includes nucleic acids that encode the GPRA polypeptides described above. The class also includes GPRA nucleic acids showing substantial sequence identity to exemplified GPRA nucleic acids (*i.e.*, SEQ ID NOS:1, 2, 4, 6, 8, 10, 12 and 14). SEQ ID NO:1 is a genomic sequence representing the part of a GPRA gene designated Ast-1 and shown in Fig. 3. SEQ ID NO:2 is a 1113 bp cDNA encoding GPRA variant A (Fig. 4A). SEQ ID NO:4 is a 1132 bp cDNA encoding GPRA variant B-Long (Fig. 4B). SEQ ID NOS:6, 8, 10, 12, and 14 encode GPRA variants B-Short, C, D, E and F.

The class also includes GPRA nucleic acids that hybridize under highly stringent conditions with at least one of the exemplified nucleic acids. Some GPRA nucleic acids of the invention hybridize under highly stringent conditions with at least one of the GPRA nucleic acids of the invention without hybridizing under the same conditions to others. For example, some GPRA nucleic acids hybridize to at least one of SEQ ID NOS:1, 4, 6, 8, 10, 12 and 14 without hybridizing to SEQ ID NO:2. Some GPRA nucleic acids include a polymorphic site occupied by a variant form as shown in Table 3 or Table 7. Some GPRA nucleic acids are genomic or minigene. Such nucleic acids contain at least one intronic sequence from any of the intron of a GPRA gene. Inclusion of an intronic sequence can be useful for increasing expression levels of a nucleic acid in cells or transgenic animals. Inclusion of an intronic sequence from introns 2, 3 or 4 is particularly useful for analyzing splice variation of the GPRA gene. Nucleic acids also include fragments of exemplified sequences SEQ ID NOS:1, 4, 6, 8, 10, 12 and 14. The fragments typically contain up to 10, 20, 25, 50, 100, 300, 400, 500 or 1000 or 1500 nucleic acids from any of the above SEQ ID NOS. Optionally, fragments include a polymorphic site shown in Table 3 or 7. The polymorphic site can be occupied by either a reference or variant form.

GPRA nucleic acids can be linked to other nucleic acids with which they are not naturally associated such as vector or a heterologous promoter sequence. Preferred nucleic acids of the invention include or are immediately adjacent to at least one of the polymorphic sites shown in Table 3 or 7. Such nucleic acids are useful as primers or probes for detection of specific alleles. Other nucleic acids are useful for expressing proteins encoded by the AST-1 locus in cells or transgenic animals. Other nucleic acids are useful for achieving gene suppression either by homologous recombination to generate a knockout animal or an antisense or siRNA

mechanism. Nucleic acids are also useful for identifying, purifying, and isolating nucleic acids encoding other, non-human, mammalian forms of asthma locus-1.

Nucleic acids of the invention can be isolated using standard molecular biology techniques and the provided sequence information (Ausubel *et al*, In Current protocols in Molecular Biology, John Wiley and Sons, publishers, 1989); Sambrook *et al*, *supra*). Cognate forms (*i.e.*, nucleic acids encoding asthma locus-1 molecules derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

V. AAA1 Nucleic Acids

The present invention provides isolated nucleic acids encoding AAA1 polypeptides. The nucleic acids can be, for example, genomic, cDNA, RNA or mini-gene (*i.e.*, a hybrid of genomic and cDNA). Genomic and mini-gene nucleic acids contain at least one intronic segment from an AAA1 gene, meaning the section of human chromosome 7p15-p14 having the structure shown in Fig. 13, or allelic and cognate variants thereof. Nucleic acids of the invention can include coding regions, intronic regions, 3' untranslated regions, 5' untranslated regions, 3' flanking regions, 5' flanking regions, enhancers, promoters and other regulatory sequences. A class of AAA1 nucleic acids can be defined with reference to exemplified AAA1 polypeptide and nucleic acid sequences. The class includes allelic, cognate and induced variants of the exemplified sequences, as well as nucleotide substitutions, which due to the degeneracy of the genetic code, do not effect the amino acid sequence of an encoded polypeptide. The class includes nucleic acids that encode the AAA1 polypeptides described above. The class also includes AAA1 nucleic acids showing substantial sequence identity to exemplified AAA1 nucleic acids (*i.e.*, SEQ ID NOS:1, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40). SEQ ID NO:1 is a genomic sequence representing the part of an AAA1 gene designated Ast-1 and shown in Fig. 13. SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 encode cDNA splice variants shown in Fig. 20.

The class also includes AAA1 nucleic acids that hybridize under highly stringent conditions with at least one of the exemplified nucleic acids. Some AAA1 nucleic acids of the invention hybridize under highly stringent conditions with at least one of the exemplified nucleic acids (*i.e.*, SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40) without hybridizing

under the same conditions to other exemplified nucleic acids. Some AAA1 nucleic acids include a polymorphic site occupied by a variant form as shown in Table 12. Some AAA1 nucleic acids are genomic or minigene. Such nucleic acids contain at least one intronic sequence from any of the introns of an AAA1 gene. Nucleic acids also include fragments of exemplified sequences SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40. The fragments typically contain up to 10, 20, 25, 50, or 100, nucleotides from any of the above SEQ ID NOS. Optionally, fragments include a polymorphic site shown in Table 12. The polymorphic site can be occupied by either a reference or variant form.

AAA1 nucleic acids can be linked to other nucleic acids with which they are not naturally associated such as vector or a heterologous promoter sequence. Preferred nucleic acids of the invention include or are immediately adjacent to at least one of the polymorphic sites shown in Table 3 or Table 12. Such nucleic acids are useful as primers or probes for detection of specific alleles. Other nucleic acids are useful for expressing AAA1 polypeptides in cells or transgenic animals. Other nucleic acids are useful for achieving gene suppression either by homologous recombination to generate a knockout animal or an antisense or siRNA mechanism. Nucleic acids are also useful for identifying, purifying, and isolating nucleic acids encoding other, non-human, mammalian forms of the AAA1 gene.

As well as or instead of encoding proteins, AAA1 transcripts can have roles in regulating expression of GPRA nucleic acids. A class of noncoding (ncRNA) has been described with regulatory activity in gene transcription and splicing (see e.g. reviews by Michel 2002; Numata *et al.* 2003).

VI. Antibodies

The GPRA and AAA1 polypeptides of the invention are useful for generating antibodies. The antibodies can be polyclonal antibodies, distinct monoclonal antibodies or pooled monoclonal antibodies with different epitopic specificities. Monoclonal antibodies are made from antigen-containing fragments of the protein by standard procedures according to the type of antibody (see, *e.g.*, Kohler, *et al.*, *Nature*, 256:495, (1975); and Harlow & Lane, *Antibodies, A Laboratory Manual* (C.S.H.P., NY, 1988) Queen *et al.*, *Proc. Natl. Acad. Sci. USA* 86:10029-10033 (1989) and WO 90/07861; Dower *et al.*, WO 91/17271 and McCafferty *et al.*, WO 92/01047 (each of which is incorporated by reference for all purposes). Phage display technology can also be used to mutagenize CDR regions of antibodies previously

shown to have affinity for the peptides of the present invention. Some antibodies bind to an epitope present in one form of GPRA or AAA1 but not others. For example, some antibodies bind to an epitope within amino acids 343-377 of GPRA long B (SEQ ID NO:4). Some antibodies specifically bind to a GPRA polypeptide containing a variant form at a polymorphic site shown in Table 7 without binding to polypeptides containing a reference form at the site. The antibodies can be purified, for example, by binding to and elution from a support to which the polypeptide or a peptide to which the antibodies were raised is bound.

Antibodies of the invention are useful, for example, in screening cDNA expression libraries and for identifying clones containing cDNA inserts which encode structurally-related, immunocrossreactive proteins. See, for example, Aruffo & Seed, *Proc. Natl. Acad. Sci. USA* 84:8573-8577 (1977) (incorporated herein by reference in its entirety for all purposes).

Antibodies are also useful to identify and/or purify immunocrossreactive proteins that are structurally related to SEQ ID NOS:3, 5, 7, 9, 11, 13 and 15 and 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 and fragments thereof used to generate the antibody.

VII. Polymorphisms of the Invention

The present invention provides a collection of polymorphic sites both single nucleotide polymorphisms (SNPs) and deletion/insertion polymorphisms (DIPs) in the GPRA gene. SNPs and DIPs can be used in mapping the human genome and, when a SNP or a DIP is linked with a disease or condition, to clarify genetic basis of the disease or condition, in this particular case, at least of asthma.

Table 3 describes 169 SNPs and 18 DIPs occurring within exons 3 and 4 and introns 2, 3 and 4 of a GPRA gene having variant forms associated with asthma. For AAA1 the corresponding polymorphisms are located in exons 4, 6, 9, 10a /b and introns 2, 3, 4, 5, 6, 7, 8, 9, and 10 (Figure 13). The locus containing these polymorphisms is also referred to as AST-1 (SEQ ID NO:1). Some of the variant forms of these polymorphism may associate with asthma due to an effect on the splicing of GPRA and/or AAA1 gene in the formation of a transcript. Other polymorphic sites may be in linkage disequilibrium with such sites or sites in coding regions that exert effects through changing the activity of GPRA or AAA1 polypeptides. Seven different haplotypes H1-H7 were identified that explain 94.5% of the genetic variation detected at the AST-1 locus and are described in Table 3. Haplotype H1 was found identical to the reference sequence NT_000380. The first column in Table 3 indicates

the types of polymorphism (*e.g.*, SNP or insertion/deletion). The second column indicates the position that the polymorphism occurs in SEQ ID NO:1. The third column indicates the position of the polymorphism occurs in NT_000380. The fourth column indicates several contiguous nucleotides flanking the polymorphic sites and the nature of the variant (*i.e.*, asthma associated) and reference/wildtype polymorphic forms. In all polymorphic sites, the allele present in SEQ ID NO:1 and NT_000380 is indicated first, and the allele from another source second. Columns 5-11 indicate alleles present at that site in haplotypes H1-H7. Column 12 indicates the location of polymorphisms in GPRA and AAA1 exons. In the AST-1 locus, there are coding polymorphism in AAA1 gene (exon 6, Asn ->Ser) and one in GPRA gene (exon 3, Asn -> Ile). Table 13 shows polymorphisms that are unique for haplotypes H1-H7, respectively, and Table 14 polymorphisms that can be employed to recognize specific haplotype combinations.

The SNPs include the following substitutions with the variant form being indicated before the reference form of SEQ ID NO: 1. Position 1 (preferably T (*i.e.*, variant form) instead of C (reference form)), position 93 (preferably G instead of A), position 918 (preferably A instead of G), position 983 (preferably T instead of A), position 987 (preferably C instead of T), position 1542 (preferably T instead of C), position 1710 (preferably G instead of A), position 1818 (preferably T instead of C), position 1927 (preferably T instead of A), position 2254 (preferably C instead of T), position 2937 (preferably A instead of G), position 3877 (preferably C instead of G), position 4012 (preferably A instead of C), position 4631 (preferably C instead of T), position 4689 (preferably G instead of C), position 4961 (preferably G instead of A), position 5442 (preferably G instead of C), position 5634 (preferably A instead of G), position 5850 (preferably G instead of A), position 6312 (preferably T instead of C), position 6392 (preferably T instead of G), position 6485 (preferably A instead of G), position 6522 (preferably G instead of C), position 6646 (preferably G instead of A), position 6739 (preferably A instead of G), position 6760 (preferably C instead of T), position 7125 (preferably A instead of C), position 7229 (preferably T instead of C), position 7277 (preferably G instead of C), position 7303 (preferably T instead of G), position 7305 (preferably C instead of G), position 7496 (preferably T instead of C), position 7550 (preferably A instead of G), position 8490 (preferably T instead of C), position 9649 (preferably G instead of T), position 10816 (preferably C instead of T), position 11858 (preferably G instead of A), position 12581 (preferably C instead of G), position 16845 (preferably C instead of T), position 16893

(preferably C instead of T), position 16980 (preferably C instead of T), position 17147
(preferably A instead of G), position 17209 (preferably A instead of C), position 17435
(preferably G instead of A), position 18383 (preferably A instead of G), position 18927
(preferably T instead of G), position 18978 (preferably G instead of A), position 19268
(preferably C instead of G), position 19272 (preferably T instead of A), position 19360
(preferably A instead of G), position 19452 (preferably A instead of G), position 19671
(preferably A instead of G), position 19712 (preferably A instead of C), position 19774
(preferably A instead of C), position 20038 (preferably C instead of T), position 20089
(preferably A instead of T), position 20309 (preferably A instead of G), position 20395
(preferably C instead of T), position 20789 (preferably T instead of G), position 21850
(preferably T instead of C), position 22475 (preferably C instead of T), position 22493
(preferably G instead of A), position 22715 (preferably A instead of G), position 22869
(preferably C instead of T), position 22934 (preferably T instead of A), position 24007
(preferably C instead of T), position 24264 (preferably T instead of G), position 24869
(preferably C instead of T), position 26198 (preferably T instead of C), position 26356
(preferably T instead of C), position 26675 (preferably G instead of A), position 27404
(preferably T instead of C), position 28197 (preferably A instead of G), position 28770
(preferably T instead of C), position 28785 (preferably A instead of G), position 28858
(preferably C instead of T), position 28866 (preferably C instead of G), position 31224
(preferably A instead of G), position 31910 (preferably A instead of G), position 32124
(preferably A instead of C), position 32185 (preferably C instead of T), position 32976
(preferably C instead of T), position 33350 (preferably A instead of C), position 33798
(preferably A instead of G), position 34362 (preferably C instead of G), position 34716
(preferably C instead of A), position 35559 (preferably T instead of C), position 36551
(preferably A instead of G), position 36909 (preferably A instead of C), position 37327
(preferably T instead of G), position 37415 (preferably A instead of G), position 37685
(preferably G instead of A), position 37931 (preferably T instead of C), position 37959
(preferably T instead of C), position 39314 (preferably A instead of G), position 39343
(preferably T instead of G), position 39927 (preferably C instead of T), position 45826
(preferably T instead of C), position 50197 (preferably T instead of G), position 50334
(preferably A instead of G), position 50493 (preferably A instead of G), position 50632
(preferably A instead of G), position 50835 (preferably C instead of A), position 50955
(preferably C instead of G), position 51217 (preferably A instead of G), position 51476
(preferably G instead of A), position 51536 (preferably T instead of C), position 51861

(preferably G instead of C), position 51884 (preferably T instead of G), position 51975 (preferably C instead of G), position 52573 (preferably A instead of G), position 52776 (preferably C instead of G), position 53803 (preferably T instead of C), position 53922 (preferably C instead of T), position 54148 (preferably A instead of G), position 54199 (preferably C instead of T), position 54641 (preferably C instead of G), position 54751 (preferably C instead of T), position 55000 (preferably A instead of G), position 55134 (preferably A instead of G), position 56683 (preferably T instead of C), position 56856 (preferably T instead of C), position 57790 (preferably C instead of A), position 60559 (preferably G instead of C), position 60604 (preferably T instead of A), position 61165 (preferably A instead of G), position 64559 (preferably T instead of C), position 65171 (preferably G instead of A), position 65857 (preferably G instead of T), position 66164 (preferably T instead of C), position 66190 (preferably T instead of C), position 66526 (preferably T instead of C), position 66902 (preferably G instead of A), position 67857 (preferably A instead of T), position 67919 (preferably C instead of T), position 72270 (preferably A instead of G), position 75115 (preferably A instead of G), position 76080 (preferably T instead of C), position 76101 (preferably C instead of G), position 81912 (preferably T instead of A), position 82203 (preferably G instead of A), position 82332 (preferably T instead of C), position 82922 (preferably A instead of T), position 83552 (preferably T instead of C), position 85227 (preferably C instead of T), position 85271 (preferably A instead of G), position 107610 (preferably C instead of T), position 110989 (preferably C instead of T), position 111012 (preferably T instead of C), position 112030 (preferably C instead of G), position 112037 (preferably G instead of T), position 112283 (preferably T instead of A), position 112726 (preferably A instead of T), position 112859 (preferably C instead of A), position 113428 (preferably G instead of C), position 113645 (preferably T instead of A), position 113944 (preferably G instead of T), position 114945 (preferably G instead of A), position 115192 (preferably G instead of C), position 115628 (preferably C instead of T), position 116032 (preferably A instead of G), position 116464 (preferably A instead of G), position 116515 (preferably A instead of G), position 116926 (preferably C instead of T), position 117276 (preferably A instead of T), position 123667 (preferably A instead of G), position (preferably 123770 instead G of A), position 123788 (preferably C instead of G), and position 129017 (preferably G instead of A).

DIPS of the invention defined relative to SEQ ID NO: 1 include: positions 184-185 (preferably insertion of AAGATA), positions 1202-1206 (preferably deletion of TAAGT),

positions 6786-6817 (preferably repeat of (TAAA)₇), position 6821 (preferably deletion of T), positions 7240-7243 (preferably deletion of ACTT), positions 7306-7308 (preferably deletion of TGT), positions 7334-7335 (preferably deletion of AT), positions 9012-9035 (preferably repeat of (CT)₁₀), positions 9199-9201 (preferably deletion of TCT), : position 9355-9356 (preferably insertion of T), positions 9782-9785 (preferably deletion of GTCT), positions 22122-22123 (preferably deletion of TG), positions 26929-26968 (preferably repeat of (TAAA)₁₁), positions 28495-28564 (preferably repeat of (CA)₁₂), position 34909 (preferably deletion of T), positions 38850-38852 (preferably deletion of CTC), positions 51022-51049 (preferably repeat of (CA)₈), and positions 52286-52287 (preferably insertion of CC).

Table 7 provide 13 additional polymorphisms occurring within exonic regions of a GPRA gene. The next eight columns indicate the position of a polymorphic site in SEQ ID NO:1, 2, 4, 6, 8, 10, 12 or 14 respectively. The next column indicates a polymorphic site and flanking nucleotides. The form of the polymorphic site present in one or more of SEQ ID NOS:1, 2, 4, 6, 8, 10, 12, and 14 is shown first and the form present in another source is indicated second. In all cases except those indicated by a *, the polymorphic form indicated first is the wildtype or reference form and the polymorphic form indicated second is the asthma or IgE associated form. In the three polymorphic sites indicated with a *, the variant form of the polymorphic site is found in NT_000380 (and SEQ ID NO:1, 2, 4, 6, 8, 10, 12 and/or 14), and the reference or wildtype allele is found in a different individual. The last column of the table shows whether an amino acid change occurs due to the polymorphisms. The positions of these polymorphisms defined with respect to SEQ ID NO:2 are: position 448 (T (variant) preferably to A (wildtype or reference), position 776 (C preferably to T), position 851 (C preferably to G), position 1159 (G preferably to A), position 1199 (T preferably to C), and position 1529 (C preferably to T); or in any one of the following positions as defined by SEQ ID NO:4: position 1206 (A preferably to C), position 1225 (T preferably to C), position 1330 (T preferably to A), and position 1338 (G preferably to A); or in any one of the following positions as defined by SEQ ID NO:8: position 585 (T preferably to C), position 655 (C preferably to A), and position 681 (A preferably to G).

Table 8 shows the five polymorphic sites within Table 7 that showed the highest association with high serum IgE level. Two of the polymorphic sites involved nonconservative amino acid changes, one within the important region of exon 3 within AST-1. Two of the

polymorphic sites were within coding regions but did not cause amino acid changes. These sites may show strong association as a result of linkage disequilibrium with causative polymorphic sites. A fifth polymorphic site is within the 3' UTR of exon 9A. This site may show a strong association for the same reason or because the nucleotide substitution has an effect on mRNA stability.

Table 12 shows associations for five polymorphic sites occurring within the AAA1 gene. At each site, the wildtype polymorphic form is indicated first, and the variant form second. The table shows relative frequency of associating haplotypes in individuals with high serum IgE level compared to those in control individuals with low serum IgE level. All haplotypes shown in the table gave chi square values of at least 8.9 ($P < 0.01$) for association.

VIII. Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by *e.g.*, PCR. See generally PCR Technology: Principles and Applications for DNA Amplification (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, Nucleic Acids Res. 19, 4967 (1991); Eckert *et al.*, PCR Methods and Applications 1, 17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, Genomics 4, 560 (1989), Landegren *et al.*, Science 241, 1077 (1988), transcription amplification (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded

RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

The identity of bases occupying the polymorphic sites shown in Table 3, 7 or 12 can be determined in an individual (*e.g.*, a patient being analyzed) by several methods, which are described in turn.

1. Single Base Extension Methods

Single base extension methods are described by *e.g.*, US 5,846,710, US 6,004,744, US 5,888,819 and US 5,856,092. In brief, the methods work by hybridizing a primer that is complementary to a target sequence such that the 3' end of the primer is immediately adjacent to but does not span a site of potential variation in the target sequence. That is, the primer comprises a subsequence from the complement of a target polynucleotide terminating at the base that is immediately adjacent and 5' to the polymorphic site. The hybridization is performed in the presence of one or more labeled nucleotides complementary to base(s) that may occupy the site of potential variation. For example, for a biallelic polymorphism two differentially labeled nucleotides can be used. For a tetraallelic polymorphism four differentially labeled nucleotides can be used. In some methods, particularly methods employing multiple differentially labeled nucleotides, the nucleotides are dideoxynucleotides. Hybridization is performed under conditions permitting primer extension if a nucleotide complementary to a base occupying the site of variation in the target sequence is present. Extension incorporates a labeled nucleotide thereby generating a labeled extended primer. If multiple differentially labeled nucleotides are used and the target is heterozygous then multiple differentially labeled extended primers can be obtained. Extended primers are detected providing an indication of which base(s) occupy the site of variation in the target polynucleotide. The methods are particularly useful for SNPs.

2. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by *e.g.*, Saiki *et al.*, Nature 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment (corresponding segments being defined by maximum alignment of the sequences being compared) from another individual

due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (*e.g.*, in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

3. Tiling Arrays

SNPs and insertion deletion polymorphism can be detected by hybridizing target nucleic acids to arrays of probes tiling a reference sequence or one or more variant forms of the reference sequence as described by WO 95/11995 (incorporated by reference in its entirety for all purposes).

4. Allele-Specific Amplification Methods

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, Nucleic Acid Res. 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. In some methods, the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. See, *e.g.*, WO 93/22456.

5. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy- chain termination method or the Maxam -Gilbert method (see Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, Recombinant DNA Laboratory Manual, (Acad. Press, 1988)).

6. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., PCR Technology, Principles and Applications for DNA Amplification, (W.H. Freeman and Co, New York, 1992), Chapter 7.

7. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, Proc. Nat. Acad. Sci. 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

8. Other Methods

Include oligonucleotide ligation assay or restriction fragment length polymorphism (RFLP) (see, for example, Current Protocols in Molecular Biology, eds. Ausubel *et al.*, John Wiley & Sons:1992, and Landegren *et al.*, "Reading Bits of Genetic Information: Methods for Single-Nucleotide Polymorphism Analysis", Genome Research 8:769-776).

VII. Methods of Use

A. Discovery of New Polymorphic Sites in AST-1 Locus

Additional polymorphic sites beyond those listed in Tables 3, 7 and 12 within GPRA and AAA1 genes and particularly, the AST-1 locus thereof can be identified by comparing a

GPRA or AAA1 nucleic acid from different individuals with a reference sequence, such as SEQ. ID NO:1, 2, 4, 6, 8, 10, 12 or 14 or SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 or 40. Preferably, the different individuals suffer or at risk of an immune disorder, such as autoimmune disease, an IgE-mediated disorder, such as asthma, chronic obstructive pulmonary disease or cancer. If a new polymorphic site (*e.g.*, SNP) is found, then the link with a specific disease can be determined by comparing the frequency of variant polymorphic forms occupying this site in populations of patients with or without a disease being analyzed (or susceptibility thereto). A polymorphic site or region may be located in any part of the locus, *e.g.*, exons, introns and promoter regions of genes.

B. Diagnostic and Detection Methods

The present invention further provides means for prognostic or diagnostic assays for determining if a subject has or is increasingly likely to develop a disease associated with the variation or dysfunction of the asthma locus of the invention. Such diseases include in addition to asthma, other IgE mediated diseases, autoimmune diseases, atopic and other immune diseases, chronic obstructive pulmonary disease and cancer. Basically, such assays comprise a detection step, wherein the presence or absence of a variant polymorphic form in the asthma locus-1 relative to SEQ ID NO:1, 2, 4, 6, 8, 10, 12, or 14 or SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 is determined in a biological sample from the subject. The detection step can be performed by any of the methods described above. Analogous assays can be performed in which variations in protein sequence relative to SEQ ID NOS:3, 5, 7, 9, 11, 13 or 15 or SEQ ID NO: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 are determined.

In particular, the present invention is directed to a method of determining the presence or absence of a variant form of a polymorphic site of the invention in a biological sample from a human for diagnosis of asthma or for assessing the predisposition of an individual to asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer. The method comprises determining the sequence of the nucleic acid of a human at one or more the above SNP or DIP positions shown in Tables 3, 7 and 12. Optionally, the sample is contacted with oligonucleotide primers so that the nucleic acid region containing the potential single nucleotide polymorphism is amplified by polymerase chain reaction prior to determining the sequence.

The polymorphic sites can be analyzed individually or in sets for diagnostic and prognostic purposes. The conclusion drawn from the analysis depends on the nature and number of polymorphic sites analyzed. Some polymorphic sites have variant polymorphic forms that are causative of disease. Detection of such a polymorphic form provides at least a strong indication of presence or susceptibility to disease. Other polymorphic sites have variant polymorphic forms that are not causative of disease but are in equilibrium dislinkage with a polymorphic form that is causative. Detection of noncausative polymorphic forms provides an indication of risk of presence or susceptibility to disease. Detection of multiple variant forms at several polymorphic sites in a GPRA and or AAA1 gene and particularly the parts of these genes within the AST-1 locus thereof provides an indication of increased risk of presence or susceptibility to disease. The results from analyzing the polymorphic sites of the invention can be combined with analysis of other loci that associate with the same disease (e.g., asthma) as polymorphic sites of the invention. Alternatively, or additionally risk of disease can be confirmed by performing conventional medical diagnostic tests of patient symptoms.

C. Clinical Trials

The polymorphisms of the invention are also useful for conducting clinical trials of drug candidates for the diseases noted above particularly asthma. Such trials are performed on treated or control populations having similar or identical polymorphic profiles at a defined collection of polymorphic sites. Use of genetically matched populations eliminates or reduces variation in treatment outcome due to genetic factors, leading to a more accurate assessment of the efficacy of a potential drug.

Furthermore, the polymorphisms of the invention may be used after the completion of a clinical trial to elucidated differences in response to a given treatment. For example, the set of polymorphisms may be used to stratify the enrolled patients into disease sub-types or classes. It may further be possible to use the polymorphisms to identify subsets of patients with similar polymorphic profiles who have unusual (high or low) response to treatment or who do not respond at all (non-responders). In this way, information about the underlying genetic factors influencing response to treatment can be used in many aspects of the development of treatment (these range from the identification of new targets, through the design of new trials to product labeling and patient targeting). Additionally, the polymorphisms may be used to identify the genetic factors involved in adverse response to treatment (adverse events). For

example, patients who show adverse response may have more similar polymorphic profiles than would be expected by chance. This allows the early identification and exclusion of such individuals from treatment. It also provides information that can be used to understand the biological causes of adverse events and to modify the treatment to avoid such outcomes.

VIII. Kits

The invention also features diagnostic or prognostic kits for use in detecting the presence of a SNP in a GPRA and/or AAA1 gene or particularly an AST-1 locus thereof in a biological sample. The kit provides means for the diagnostics of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer or for assessing the predisposition of an individual to a disease mediated by variation or dysfunction of a GPRA and/or AAA1 gene. The kit can comprise a labeled compound capable of detecting the nucleic acid of a GPRA and/or AAA1 gene in a biological sample. The kit can also comprise nucleic acid primers or probes capable of hybridizing specifically to at least a portion of a GPRA gene. Preferably, the primer is a minisequencing primer specific to any one of the above polymorphisms. The kit can be packaged in a suitable container and preferably it contains instructions for using the kit.

Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 50 or all of the polymorphisms shown in Tables 3, 7, 12, 13, and 14. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions.

IX. Transgenic Animals and Cells

The present invention also makes available plasmids and vectors comprising the nucleic acid of the invention, which constructs can be used when studying biological activities of the AST-1 locus. The selecting of a suitable plasmid or vector for a certain use is within the abilities of a skilled artisan. As the host cell may be any prokaryotic or eukaryotic cell, a plasmid or vector comprising the locus can be used in the duplication of the locus (see, for

example, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", Second Edition, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (1989)).

Nucleic acids of the invention can be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful reagents. A transgenic animal (*e.g.*, a bovine, pig, sheep, rabbit, rat, mouse or other rodent) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at an embryonic (*e.g.*, one-cell) stage. Often all or substantially all of the somatic and germline cells of the transgenic animal have a copy of the transgene in their genome. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. Methods for generating transgenic animals, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009. Typically, a GPRA and/or AAA1 nucleic acid is operably linked to a promoter and optionally an enhancer. The promoter and enhancer can be selected confer tissue specific expression. Transgenic animals that include a copy of a transgene introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of GPRA and/or AAA1 polypeptides. Such animals can be used as tester animals for reagents thought to confer protection from, for example, diseases related to GPRA and/or AAA1. Such a transgenic animal can be treated with agent and a reduced incidence of a characteristic of disease compared to untreated animals (*i.e.*, without the agent) bearing the transgene indicates a potential therapeutic intervention for the disease. Results from animal models can be confirmed in placebo-controlled clinical trials on patients with asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

Transgenic offspring are identified by demonstrating incorporation of the microinjected transgene into their genomes, preferably by preparing DNA from short sections of tail and analyzing by Southern blotting for presence of the transgene ("Tail Blots"). A preferred probe is a segment of a transgene fusion construct that is uniquely present in the transgene and not in the mouse genome. Alternatively, substitution of a natural sequence of codons in the transgene with a different sequence that still encodes the same peptide yields a unique region identifiable in DNA and RNA analysis. Transgenic "founder" mice identified in this fashion are bred with normal mice to yield heterozygotes, which are backcrossed to create a line of transgenic mice. Tail blots of each mouse from each generation are examined until the strain

is established and homozygous. Each successfully created founder mouse and its strain vary from other strains in the location and copy number of transgenes inserted into the mouse genome, and hence have widely varying levels of transgene expression or activity.

Knockout animals or cells can also be made by functionally disrupting an endogenous cognate form of one or more genes within the AST-1 locus (i.e., GPRA and/or AAA1). Functional disruption means that the transgenic animal is incapable of making a functional GPRA or AAA1 polypeptide and/or transcript encoding the same. Inactivation of a protein can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker within the coding region. See Capecchi, Science 244, 1288-1292 (1989). Inactivation of a transcript can be achieved by mutation within a promoter region. The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

X. Methods of Screening for Drugs

The GPRA and AAA1 polypeptides and nucleic acids of the invention of the invention are useful for screening for agents that modulate (e.g., agonize or antagonize) their function (e.g., as a G-protein coupled receptor). Such agents can be useful as drugs in compensating for genetic variations affecting GPRA and/or AAA1 function. For example, an agonist of GPRA and/or AAA1 function is useful in a patient having a genetic variation that decreases endogenous GPRA function, and an antagonist is useful in a patient having a genetic variation that increases endogenous GPRA and/or AAA1 function. GPRA and AAA1 polypeptides can also be used to screen known drugs to determine whether the drug has an incidental effect on GPRA and/or AAA1 activity. A drug that has an incidental agonist activity should generally be avoided in patients having atypically high levels of IgE antibodies, and an antagonist should generally be avoided in immunosuppressed patients.

Agents can initially be screened in a binding assay with a GPRA or AAA1 polypeptides. A binding assay reduces the pool of initial agents to a subset that can be screened by a functional assay. Agents can be screened in cells transfected with nucleic acids encoding a GPRA and/or AAA1 polypeptide of the invention or in transgenic animals transfected with a GPRA and/or AAA1 nucleic acids and optionally and a reporter construct containing cAMP response element, a minimal promoter and a coding sequence for secreted alkaline

phosphatase. Methods for identifying agents that have a functional activity in transducing a signal through a cellular receptor are described by US 6,309,842. Agents for screening can be obtained by producing and screening large combinatorial libraries, as described above, or can be known drugs, or can be known ligands to related receptors to GPRA.

Agents can also be screened for capacity to inhibit or increase expression of GPRA and/or AAA1 nucleic acids. Suitable agents to screen include antisense polynucleotides, see *Antisense RNA and DNA*, (1988), D.A. Melton, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY); Dagle et al., *Nucleic Acids Research*, 19:1805 (1991); Uhlmann et al., *Chem. Reviews*, 90:543-584 (1990), zinc finger proteins (see WO 00/00388 and EP.95908614.1), and siRNAs WO 99/32619, Elbashir, *EMBO J.* 20, 6877-6888 (2001) and Nykanen *et al.*, *Cell* 107, 309-321 (2001); WO 01/29058.

Agents having pharmacological activity evidenced by cellular or animal testing can be subject to clinical trials in patients suffering from asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer in comparison with a placebo.

XI. Methods of Treatment

The invention provides methods of treating asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer. These methods entail administering an effective amount of a GPRA and/or AAA1 polypeptide or an agent, e.g., an antibody, that modulates activity of a GPRA and/or AAA1 polypeptide. The agent can modulate that activity for example, by an effect on signal transduction through the GPRA polypeptide (usually a decrease) or by effecting expression of the GPRA and/or AAA1 polypeptide (also usually a decrease). An agent can be administered for prophylactic and/or therapeutic treatments. A therapeutic amount is an amount sufficient to remedy a disease state or symptoms, or otherwise prevent, hinder, retard, or reverse the progression of disease or any other undesirable symptoms in any way whatsoever. In prophylactic applications, an agent is administered to a patient susceptible to or otherwise at risk of a particular disease or infection. Hence, a "prophylactically effective" amount is an amount sufficient to prevent, hinder or retard a disease state or its symptoms. In either instance, the precise amount of compound contained in the composition depends on the patient's state of health and weight.

An appropriate dosage of the pharmaceutical composition is determined, for example, using animal studies (*e.g.*, mice, rats) are commonly used to determine the maximal tolerable dose of the bioactive agent per kilogram of weight. In general, at least one of the animal species tested is mammalian. The results from the animal studies can be extrapolated to determine doses for use in other species, such as humans for example.

The pharmaceutical compositions can be administered in a variety of different ways. Examples include administering a composition containing a pharmaceutically acceptable carrier via oral, intranasal, rectal, topical, intraperitoneal, intravenous, intramuscular, subcutaneous, subdermal, transdermal, intrathecal, and intracranial methods. The route of administration depends in part on the chemical composition of the active compound and any carriers.

The components of pharmaceutical compositions are preferably of high purity and are substantially free of potentially harmful contaminants (*e.g.*, at least National Food (NF) grade, generally at least analytical grade, and more typically at least pharmaceutical grade). To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions for parental administration are also sterile, substantially isotonic and made under GMP conditions. Compositions for oral administration need not be sterile or substantially isotonic but are usually made under GMP conditions.

MATERIALS AND METHODS

Families were recruited in central eastern Finland in 1994 and 1996. The methods for recruitment, control for population stratification, and clinical evaluation have been described previously in detail (Kauppi *et al.* 1998; Laitinen *et al.* 1997). We used self-reported asthma as a sampling method. Altogether 253 families were recruited, two thirds of which were trios. Based on retrospective verification of the disease history and the results of diagnostic tests (spirometry, histamine or methacholine challenge test, expiratory peak flow measurements), 87% of the self-reported asthma patients were accepted as verified cases (Kauppi *et al.* 1998). Criteria for asthma were based on the recommendations of the American Thoracic Society (Dantzker, D.R. *et al.* 1987). In addition to the 86 multiplex pedigrees included already into

our previous genome scan, 103 nuclear families with full phase information (a total of 874 study individuals) were included to the association analysis without further selection.

Total serum IgE level was determined by Diagnostics CAP FEIA (Kabi Pharmacia, Sweden) in one batch. Allergy screening was done to all study individuals at the time of the study (Kauppi *et al.* 1998; Laitinen *et al.* 1997). Based on total serum IgE level the study individuals were divided into two groups: high IgE responders ($\text{IgE} > 100\text{kU/L}$) and low IgE responders ($\text{IgE} \leq 100\text{kU/L}$).

EXAMPLES

EXAMPLE I

Construction Of The Physical Map

We have ordered the markers used in the analysis and estimated the physical distance between the markers using several publicly available sources as described previously by Polvi *et al.* (2002). Completed chromosome 7 sequence confirms the distance and order of our markers

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Nucleotide&list_uids=22096428&dopt=GenBank).

EXAMPLE II

Microsatellite Discovery and Genotyping

To create a dense map of polymorphic markers spread evenly across the linkage region, we screened the publicly available genomic sequences for potentially polymorphic tandem repeats (Polvi *et al.* 2002). Eighty microsatellites were genotyped in the whole data set (874 samples) giving the average marker density of 220kb (range from 904bp to 790kb) across the linkage region. All microsatellite markers and small insertions and deletions (Table 1C) found in the critical region were genotyped using fluorescently labeled primers in gel electrophoresis on the ABI377 sequencer as described previously (Polvi *et al.* 2002 and <http://www.genome.helsinki.fi/eng/research/asthma/detail.htm>).

EXAMPLE III

Discovery of Genetic Variation in the Critical

Region by Direct Sequencing

To identify all sequence polymorphisms in the critical region, we generated a 129 kb reference sequence by assembling the sequence (genomic clones AC005680, AC005492, AC005174, AC005862.1, and AC005853.1) from the Human Genome Project (in the contig NT_000380 from the position 509,783 to the position 638,799). SNP discovery by re-sequencing was done in four selected patients and in three control individuals who were homozygous across the region and for the flanking markers. Repeat regions such as SINE, LINE, LTR, MER1 and MER2 elements covered 60% of the sequence (a total of 129,017 bp). PCR assays were designed to be ~650 bp in length with 100 bp overlap with adjacent assays. Repeat regions were re-sequenced when it was possible to design primers in unique sequence within the range mentioned above.

PCR assays were carried out in 20µl volumes containing 20ng of genomic DNA, 1x PCR Buffer II, 0.1mM dNTPs, 2.5 mM MgCl₂, 0.1µM primer mix, and 0.5 U of DNA polymerase (AmpliTaqGold, Applied Biosystems). The samples were denatured for 10 min at 94°C, followed by 35 cycles each of 30s at 94 C, 30 s at 58 C, and 30 s at 72 °C. Elongation was performed for 10 min at 72°C. Purified PCR fragments (Quickstep 2 PCR purification Kit, Edge BioSystems, MD) were sequenced in both directions using ABI Prism3100 sequencer and dye-terminator chemistry. We assembled the sequence reads using the Gap4 program.

EXAMPLE IV

SNP Genotyping

SNP genotyping was done using two different methods: single base pair extension (SBE) or using altered restriction sites. SBE was done using the chemistry of Molecular Dynamics according to the suggestion made by the manufacturer on a Megabase 1000 sequencer (Molecular Dynamics, CA) using the primers listed in the Table 1A. Allele calling was performed by using the MegaBACE SNP Profiler software (Molecular Dynamics).

Twenty-nine of the SNPs were genotyped using different restriction enzyme digestions. All primers, restriction enzymes, and lengths of the digestion fragments of corresponding allele used in the genotyping are given in Table 1B. If the SNP did not produce a natural site for

altered restriction, mutations were induced in PCR-primers (shown in capital letters in primer sequence). To improve allele calling by growing size difference between alleles, in some primers plasmid sequence tail of 20-30bp was added (shown in bold in the primer sequence). Altered restriction sites of the PCR products stained with ethidium bromide and visualized on agarose gels in UV light were called manually by two independent observers. All the markers were in Hardy-Weinberg equilibrium and observed Mendel errors were less than 0.1%.

EXAMPLE V

Haplotype Association Analysis

The haplotype analysis was done using Haplotype Pattern Mining (HPM) (Toivonen *et al.* 2000) program. For haplotyping, large pedigrees were divided into trios using an in-house computer program. The program identifies the maximum number of trios that are not overlapping and in which one or two members were affected (not both parents). Trios that included members who had not been genotyped or members with unknown phenotype were excluded. In phase three of hierarchical LD mapping, high density genotyping of the critical region was done in the trios informative for the association analysis (a total of 132 trios, 396 study individuals, and revealing 304 unrelated affected and 220 control chromosomes).

Haplotyping was done within each trio and four independent chromosomes were obtained from each trio. In case of ambiguities (missing genotype data, identical heterozygotic genotypes in all of the family members, or Mendel errors), the alleles were discarded. If the child was affected, the transmitted chromosomes were considered disease associated and the non-transmitted chromosomes as controls. If one of the parents was affected, his/her chromosomes were considered disease associated and the spouse's chromosomes as controls. If both the parent and the child were affected, only the non-transmitted chromosome of an unaffected parent was considered as the control and the other three as disease associated. These haplotypes were used as input for HPM.

To estimate the overall significance of detected association, accounting for simultaneous testing of multiple markers, we used a second level permutation test by performing nested permutation tests described in detail elsewhere (Sevon *et al.* 2001a and b).

RESULTS

1. Haplotype Association

Our genotyping effort searching for a shared haplotype among individuals who have high total IgE serum level was done in three phases using a hierarchical approach. Based on the initial observation on potential association, the marker map was made denser in the regions of primary interest. In the haplotype analysis we used the HPM algorithm that is known to be powerful in locating a disease-causing gene when one is known to exist (Toivonen *et al.* 2000). By allowing gaps in the haplotype patterns, the algorithm is robust for genotyping errors, marker mutations, unrecognized recombinations and missing data. The data set consisted of 86 multiplex pedigrees that were included also in the previous linkage study and additional 103 nuclear families recruited solely for the association study. These families revealed 304 unrelated high IgE associated and 220 control chromosomes which were then used as the input data for the HPM analysis.

First we built a marker map across the whole linkage region with 80 microsatellite markers (Fig. 1). In the first phase of haplotype association analysis the marker density was approximately 220kb across the linkage region. The best associations (chi square ≥ 7.4) formed only one cluster of haplotypes located between the markers D7S690–NM13 (~3.21cM). In the second phase we added 6 new microsatellites in between D7S690 and NM13. The associated haplotypes (chi square ≥ 8.9) clustered now at the centromeric end of the region, between markers G42099 and NM13 (301kb) (Fig. 2). HPM algorithm was used to evaluate the overall haplotype distribution among the high IgE associated compared to that among control chromosomes. A permutation test incorporated to the HPM computer package was used to evaluate the statistical significance of our finding by comparing the overall haplotype distribution among the high IgE associated and control chromosomes. Using chi square 7.0 as the threshold for association, 7 markers as the maximum length of the haplotype, and allowing one gap in the haplotype patterns, the best associations were computed for three adjacent markers G42102, D7S497, and G4296 (marker-wise P values 0.05-0.06 based on 10 000 simulations).

Fine mapping was then further continued by adding 49 new markers between G42099 and NM13 giving an average marker density of 6kb. By using the HPM algorithm, all associated haplotype (chi square >13) patterns spanned between SNP509783 and SNP638799 (129,017 bp).

To study if the overall haplotype distribution in the locus differs in high IgE associated compared to that in control chromosomes, we used again a permutation test. HPM analysis was done using the following parameters: maximum pattern length 40 markers, one gap allowed for missing data and possible errors, and chi-square threshold for the association 13.0. The observed scores (=number of qualified haplotype patterns spanning across the marker) for associated haplotypes varied from 0 to 40. Permutation test showed statistically biased haplotype distribution for high IgE. The marker-wise P values ≤ 0.005 were observed for all eleven of the markers between SNP531632 and SNP563930 and the marker wise P values ≤ 0.01 for all twenty-eight markers between NM51 and SNP617392 based on 10 000 simulations. To estimate the effect of simultaneous testing of multiple markers, we used a second level permutation test, yielding a corrected P value of 0.03 (Sevon 2001a and b).

2. Genetic Variation in the Critical Region

Based on haplotype analysis we had determined the critical region in between the markers SNP509783 and SNP638799 (Fig. 2). SNP genotyping resulted in the identification of seven haplotypes H1-H7 that together explain 94.5% of the genetic variation detected at the AST-1 locus. To identify all potentially causative genetic variations on the risk haplotype, we sequenced four patients who carried haplotypes H2/H2, H4/H4, H5/H5, and H7/H7, respectively, and were homozygous across the AST1-locus region. The patients' sequences polymorphisms were compared with the reference sequence (SEQ ID NO:1) (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=nucleotide&list_uids=7382309&dopt=GenBank, human genomic contig NT_000380) and to three population based controls homozygous for the same markers, but with different haplotypes H1/H1, H3/H3, and H6/H6, respectively. The differences are highlighted via SEQ ID NO:1. Genetic variation including small deletions and insertions (≤ 6 base pairs), variation in the length of microsatellite repeats, and multiple SNPs are presented in Table 3 that can be used as genetic markers of the risk haplotypes. Tables 13 and 14 show the SNPs unique for each haplotype and unique for certain haplotype combinations, respectively. In diagnostic testing these markers identify the haplotypes and the haplotype combinations of an individual without phase information.

Example VI: Identification of the GPRA Gene

A. Materials and Methods

1. Exon Prediction

Exon predictions were performed by GENSCAN (<http://genes.mit.edu/GENSCAN.html>) and software using genomic clones from AC005493.1 to AC005826.1 (690 kb) (Polvi *et al.* 2002) as one block. Predictions were performed for both unmasked and masked sequence. In the latter, the repetitive sequences were masked using the Repeat Masker Service (<http://repeatmasker.genome.washington.edu/>).

2. Reverse Transcriptase PCR

The exon-specific primers were designed based on exon predictions (Table 4, Fig. 3). First strand cDNAs were synthesized from 1 µg of commercially available poly A⁺ RNAs from human brain (Invitrogen), lung, testis, placenta, and thymus (BD Clontech) by the SMART RACE cDNA Amplification Kit (BD Clontech, Palo Alto, CA) according to manufacturer's instructions. Alternatively, Marathon-Ready cDNA (Clontech) or cDNA from Multiple Tissue cDNA Panels I and II (Clontech) were used as templates in some PCR amplifications. All the PCR amplifications were performed in 20-50 µl volumes using 2.0-5.0 µl cDNA as template, 2 mM MgCl₂, 0.2 mM dNTPs (Finnzymes), 0.2 mM of each primer, 0.15-0.4 U AmpliTaqGold, and 1x PCR Gold buffer (Applied Biosciences) under the following conditions: 94°C for 10 min, 35-40 cycles of 94 °C for 30 s, 58 °C –68 °C for 30 s, 72 °C for 1 min 40 s to 2 min followed by 72 °C for 10 min. PCR products were analyzed on 1 % agarose gel and extracted from the gel using QIAquick Gel Extraction Kit (Qiagen).

3. Rapid Amplification of cDNA Ends (RACE-PCR)

To generate 3' and 5' cDNA ends, rapid amplification of cDNA ends (SMART RACE) was performed using human testis cDNA and Human Marathon-Ready Fetal Thymus cDNA (Clontech) according to the manufacturer's protocol for the SMART RACE cDNA Amplification Kit (BD Clontech). RACE-PCR products were cloned using pGEM-T Easy Vector system (Promega) or TOPO TA Cloning kit (Invitrogen) and plasmid DNA was purified using QIAprep Spin Miniprep Kit (Qiagen).

The purified RT-PCR products and the cloned RACE-PCR products were verified by automated sequencing with dye-terminator chemistry (ABI Prism3100, Applied Biosystems, Inc).

4. Cloning of the Full-Length cDNAs of GPRA

Nested PCR amplification was used in cloning of the full length cDNAs for *GPRA* splice variants A and B. The first round of PCR for variant A was done with the primer pair JEGE1F1 and JEGE9aR1, and for variant B with JEGE1F1 and JEGE9bR1 (Table 4, Fig. 3). The following inner primers were used in the re-amplification: JEGE1F2 and JEGE9aR2 for variant A; JEGE1F2 and JEGE9bR2 for variant B, respectively. Full length C variant was amplified with the primer pair JEGE1F2 and JEGE_xR1.

Primary PCR amplifications of variants A and B were performed in 25 µl volumes using 2.5 µl Human Brain Marathon-Ready cDNA (Clontech) as template, 1x DyNAzyme EXT buffer (containing 1.5 mM MgCl₂), 0.2 mM dNTPs (Finnzymes), 0.52 µM of each primer, 5% DMSO and 0.5 U DyNAzyme EXT polymerase (Finnzymes) under the following conditions: 94°C for 4 min, 38 cycles of 94°C for 30 s, 65°C for 30 s, 72°C for 1 min followed by final extension of 72°C for 10 min. The aliquot of primary PCR product was re-amplified by 30 PCR cycles under the same conditions as above. PCR amplification of variant C was performed in the above conditions. NCI-H358 cDNA was used as template.

PCR products were cloned into pCR 2.1 TOPO -vector using TOPO TA cloning kit (Invitrogen) according to manufacturer's instructions and plasmid DNA were purified using QIAprep Spin Miniprep Kit (Qiagen). The cloned RT-PCR products were verified by automated sequencing with dye-terminator chemistry (MegaBACE 1000, Molecular Dynamics).

5. Cell Culture and Isolation of Poly A⁺ RNA

In addition to commercially available poly A⁺ RNAs and Marathon-Ready cDNAs, human lung epithelial carcinoma cell line NCI-H358 (ATCC) was cultured as mRNA source for expression studies. Cells were cultured in RPMI 1640 medium (Gibco BRL) supplemented with 1 mM sodium pyruvate (Gibco BRL), 10% FCS (Biological Industries), and 1% Penicillin/Streptomycin (GibcoBRL). Poly A⁺ RNA was isolated by Dynabeads mRNA DIRECT Kit (Dyna) according to the manufacturer's instructions.

6. Northern Blot Hybridization

The GPRA C specific (comprising the nucleotides from 58 to 527 of the *GPRA* C variant) 470 bp probe was generated by RT-PCR using human testis cDNA as template. The probe was radiolabelled with α [32 P]-dCTP using RediPrime Kit (Amersham Biosciences) according to the manufacturer's instructions. A Human Multiple Tissue 8-lane Northern blot (BD Clontech) was prehybridized in ExpressHyb solution (BD Clontech) for 1 h at 68 °C followed by hybridization with the specific probe for 1 h 15 min at 68 °C. Herring sperm DNA (100 μ g/ml) was used as the blocking reagent. The filter was washed with 2 x SSC and 0.05 % SDS at room temperature and then exposed to X-ray film at -20 °C for 6 days.

7. Western Blot and Immunohistochemistry

Two specific antibodies against the alternative carboxy terminals of *GPRA* were raised by immunizing rabbits with the following peptides: CREQRSQDSRMTFRERTER (corresponding to the nucleotides 1148-1205 of the variant A) and CPQRENWKGTWPGVPSWALPR (corresponding to the nucleotides 1196-1259 of the variant B of *GPRA*). *GPRA* A peptide synthesis and antibody production were purchased from Sigma-Genosys Ltd (London Road, Pampisford, Cambridge). *GPRA* B peptide synthesis was purchased from the University of Helsinki and antibody production from the University of Oulu. A total of 6 immunizations were performed at 2 weeks intervals using the total of 2 mg of KLH-conjugated peptide purified by gel filtration. *GPRA* antibodies were purified from whole serum by affinity chromatography with the peptide coupled to iodoacetyl on a crosslinked agarose support according to the manufacturer's (Pierce, Meridian Road, Rockford IL, USA) instructions.

For western blot analysis, human tissue lysates from spleen, skeletal muscle, uterine muscle, colon muscle, kidney, colon epithelium, testes, and prostate were obtained by mechanically homogenizing the frozen tissue samples in 10 mM Tris HCl 100mM NaCl 2 % Triton X-100 buffer with proteinase inhibitors. 50 μ g of the protein lysates were run on reducing 12.5 % SDS-PAGE gels and electroblotted to the PVDF membrane according to standard procedures. Nonspecific protein binding was prevented by incubating the membrane with 5 % milk in 0.1 % Tween 20 in TBS (TBST). Antigenic sites were revealed by incubating the membrane with the anti-*GPRA* C-terminal antibodies or pre-immune serum followed by the alkaline phosphatase-conjugated goat anti-rabbit secondary antibody (Jackson ImmunoResearch Laboratories Inc., West Baltimore Pike West Grove, PA, USA) in 5 % milk in 0.1 % TBST.

The color reaction was revealed by the NBT/BCIP method (Pierce). Negative controls did not show specific reactivity.

Formalin fixed, paraffin-embedded specimens of normal adult human bronchus, skin and colon, and human normal tissue array slides (MaxArray, Zymed Laboratories Inc., CA) containing 30 different tissues were used for immunohistochemistry. For pre-treatment, slides were deparaffinized by xylene-treatment followed by decreasing alcohol series. The slides were heated in microwave oven in 10 mM citrate buffer, pH 6.0 for 5 minutes.

Immunohistochemical analyses were performed using the ABC method (Vectastain Elite ABC kit, Vector Laboratories, Burlingame, California). Omission of primary antibody and staining with the nonimmunized sera was used as negative control for parallel sections. Neither of these controls showed any immunoreactivity.

8. Transient transfections and Elisa

Cos-1 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum, 1% PS and 5% non-essential amino acids. Cos-1 cells were transiently transfected with pCMV-myc- GPRA A, B, C, D, E, F or Bshort separately using Fugene6 transfection reagent (Roche) following the manufacturer's protocol. After 24 hours, the transfected cells from one well of 6 well plate were divided into 16 wells of 96 well plate. After 48 hours of transfection, the cells were fixed with 3.5 % PFA. Half of the cells were permeabilized with 0.5% TX-100 for 10-15 min, blocked with TBS containing 2% milk and 1% goat normal serum at 37 °C for 30 min, incubated with 1:1000 dilution of myc- specific primary antibody (Babco) for 1 h at 37 °C, washed three times with TBS and thereafter incubated with a dilution of 1:2000 of HRP-conjugated anti-mouse IgG antibody for 30 min at room temperature and washed three times with TBS. TMB-substrate (Sigma) was added to cells and the reaction was let to proceed for 3 to 6 minutes after which the reaction was stopped by adding 1.5 M HCL. Absorbance was measured at 450 nm.

9. Computational Protein Sequence Analysis

The transmembrane topology and putative N-glycosylation sites of *GPRA* protein were predicted using PredictProtein (http://cubic.bioc.columbia.edu/predictprotein/submit_def.html#top) and TMPred (http://www.ch.embnet.org/software/TMPRED_form.html) softwares. In protein homology comparisons the BlastP

(<http://www.ncbi.nlm.nih.gov/blast/>) and T-Coffee (<http://www.ch.embnet.org/software/TCoffee.html>) softwares were used.

10. Family Collection

Genotyping of *GPR4* was done in a family collection recruited in central eastern Finland. The methods for recruitment, control for population stratification, and clinical evaluation have been described previously in detail (Kauppi *et al.* 1998; Laitinen *et al.* 1997). We used self-reported asthma as a sampling method. Total serum IgE level was determined by Diagnostics CAP FEIA (Kabi Pharmacia, Sweden) in one batch for all the participants. Based on total serum IgE level the study individuals were divided into two groups: high IgE responders (IgE > 100kU/L) and low IgE responders (IgE ≤ 100 kU/L).

Altogether 253 families were recruited, two thirds of which were trios. Based on retrospective verification of the disease history and the results of diagnostic tests (spirometry, histamine or methacholine challenge test, expiratory peak flow measurements), 87% of the self-reported asthma patients were accepted as verified cases (Kauppi *et al.* 1998). Criteria for asthma were based on the recommendations of the American Thoracic Society (Dantzker, D.R. *et al.* 1987). 86 large pedigrees were included into our previous genome scan (Laitinen *et al.* 2001). For the association study, those pedigrees were divided into trios using an in-house computer program. Non-overlapping trios with full phase and phenotype information and additional 103 trios were included to the association analysis without further selection.

11. Screening for Exon Polymorphisms

Genomic DNA of four asthma patients and one healthy control from above mentioned study group were screened for sequence variation with 12 primer pairs covering all the verified exons of *GPR4* (Table 5). All the patients were homozygous for microsatellites G42099, G42100, D7S497, and G42097 (from the NT_000380 position 476,213 to 644,425, a total of 168 kb) and expressed the susceptibility haplotype (Patent 1). The control individual showed homozygosity for the same markers, but to a different, non-associating haplotype.

PCR assays were carried out in 20 µl volumes containing 20 ng of genomic DNA, 0.1mM dNTPs (Finnzymes), 2.5 mM MgCl₂ (Applied Biosystems), 0.1µM primer mix, and 0.5 U of DNA polymerase, and 1x PCR Buffer II (AmpliTaQGold, Applied Biosystems). The samples

were denatured for 10 min at 94 °C, followed by 35 cycles each of 30s at 94 °C, 30 s at 58 °C, and 30 s at 72 °C. Elongation was performed for 10 min at 72 °C. Purified PCR fragments (Quickstep 2 PCR purification Kit, Edge BioSystems) were sequenced from both directions using ABI Prim3100 (Applied Biosystems) sequencer and dye-terminator chemistry. We assembled forward and reverse sequence reads using the Gap4 program (Staden Package software).

12. Genotyping

Genotyping of five exonic SNPs in *GPRA* was done using two different methods. One marker was genotyped using single base pair extension (SBE) with the chemistry of Molecular Dynamics on a Megabase 1000 sequencer (Molecular Dynamics) according to the suggestion made by the manufacturer (Table 6A). Allele calling was performed by using the MegaBACE SNP Profiler software (Molecular Dynamics).

Four of the SNPs were genotyped using different restriction enzyme digestions. All primers, restriction enzymes, and lengths of the digestion fragments of corresponding allele used in the genotyping are given in Table 6B. If the SNP did not produce a natural site for altered restriction, mutations were induced in PCR-primers. To improve allele calling by growing size difference between alleles, in one primer plasmid sequence tail was added. Altered restriction sites of the PCR products were visualized on ethidium bromide stained agarose gels in UV light and called manually by two independent observers. All the markers were in Hardy-Weinberg equilibrium and observed Mendel errors were less than 0.1%.

13. Haplotype Association Analysis

The haplotype analysis was done using Haplotype Pattern Mining (HPM) (Toivonen *et al.* 2000) program. The data set provided a total of 132 informative trios, 396 study individuals, and revealing 304 unrelated affected and 220 control chromosomes. Haplotyping was done within each trio and four independent chromosomes were obtained from each trio. In case of ambiguities (missing genotype data, identical heterozygotic genotypes in all of the family members, or Mendel errors), the alleles were discarded. If the child was affected, the transmitted chromosomes were considered disease associated and the non-transmitted chromosomes as controls. If one of the parents was affected, his/her chromosomes were considered disease associated and the spouse's chromosomes as controls. If both the parent

and the child were affected, only the non-transmitted chromosome of an unaffected parent was considered as the control and the other three as disease associated. These haplotypes were used as input for HPM.

RESULTS

1. Characterization GPRA at RNA And Protein Level

A. Genomic Structure of Different Splice Variants of GPRA

The full-length cDNA sequence of *GPRA* was assembled by using RT-PCR with primers designed for predicted exons and by using RACE-PCR to generate the 5' and 3' ends of the gene. The genomic location of the primers is shown in the Fig. 3 and sequence in the Table 4. Nested PCR amplification was used to produce full length cDNAs of different splice variants in one fragment. Using different (Marathon-Ready Brain cDNAs, RNA extracted from NCI-H358 cell line) template three splice variants were identified repeatedly (A, B_{long}, and C variants in the Fig. 3 and 4). All variants possess the same initiation site, but used alternative exons to encode the 3' end of the gene (exons 9A, 9B and 2B, respectively). The sequence flanking the putative ATG translation initiation site (GCCATGC) contains the -3 purine, but not the +4 guanine residue, of the Kozak consensus sequence. In two of the variants, the open reading frame was distributed across 9 exons producing cDNAs of 1116 bp and 1134 bp in size (variants A and B_{long}, Fig. 5A and 5B) and encoding 371 and 377 amino-acid proteins, respectively. The C variant is a shorter transcript of *GPRA* including only the exons E1, E2a and E2b (Fig. 5C). *GPRA* spanned 0.2 Mb of the genomic contig NT_000380. The second intron is the largest, comprising of 93.8 kb.

When nested PCR specific for the B variant was applied in the cloning process a new splice variant was observed (B_{short} in Figs. 4B and 5B). A 33 bp deletion was observed in the 5' prime end of the exon 3.

B. Expression Profiling of GPRA Using RT-PCR and Northern Blot Hybridization

RT-PCR analysis showed that different variants of *GPRA* were expressed in several human tissues. These tissues include testis, brain, pituitary gland, placenta, lung, heart, fetal thymus, and fetal heart. In addition to commercially available cDNAs, we analyzed the *GPRA* expression profile in NCI-H358 cell line that represents broncho-epithelial origin. Variants A (JEG5F1 and Vau8R2), B (JEG5F1 and Vau1000R1) and C (JEG5F1 and JEGEXR1)

specific primer pairs were used in amplifications (Fig. 8, Table 4). All forms were expressed and using A variant specific primers, we found several transcripts. PCR fragments were cloned and sequenced. Sequence verification revealed variants A, D, E, and F (Figs. 4 and 5). Northern blot hybridization was done using a 470 bp cDNA probe comprising exons E1, E2A, and E2B (Fig. 3). On Human 8-Lane MTN blot, the probe detected multiple transcripts sized approximately 6.5 kb, 6.0 kb, 1.8 kb, and 1.0 kb as shown for placental tissue in Fig. 9.

C. The Predicted Structure of the GPRA Protein

Both TMpred and PredictProtein softwares predicted that the variants A and B_{long} encode a 7TM protein with an extracellular N-terminus (approximately 50 amino acids) and intracellular C-terminus (Figs. 6A-6F). Their structure shared the common features of GPR family A with 16 conserved amino acids in the TM, loop, and C-terminal regions. These include among others two conserved Cys residues in exoloops 1 and 2 that potentially form a disulfide bridge; the Asp-Arg-Tyr sequence (DRY motif) in the proximity of TM3; Asn-Pro-X-X-Tyr motif in TM7; and a Cys residue in the C-terminal region (Figs. 6A-6F). The putative N-glycosylation site Asn⁴ was predicted by PredictProtein. All the other variants lacked the 7TM structure. For variant B_{short} a 6TM structure was predicted (Fig. 6B). Variant C encodes 94 amino acid protein that has potentially only one TM region (Fig. 6C). For the variants D and E, deletions of the exons 3 or 4 caused a shift in the ORF, presumably producing truncated forms of the protein. When both exons were deleted (variant F), the ORF remained intact, but only five-transmembrane regions were predicted (Fig. 6F).

D. Sequence Searches

Determined by BlastP comparisons, the GPRA protein shows 31 % amino acid identity to human vasopressin receptor 1B, 28 % identity to human vasopressin receptor 2, 32 % identity to human oxytocin receptor, and 43 % amino acid sequence identity to gene product CG6111 of *Drosophila melanogaster* which is considered as an orthology to human vasopressin/oxytocin receptor family (Park *et al.* 2002). The best sequence identity between proteins was found in TM regions and in the first exoloop. Based on T Coffee homology comparisons, GPRA showed 20% amino acid identity to the bovine rhodopsin receptor. Sequence comparisons between GPRA and mouse ESTs BB638128, BB632343, BB625809, BB228269, and publicly available mouse genomic sequence suggest the existence of a mouse orthology.

2. Expression Profiling of GPRA Variants A and B Using Western Blot Analysis and Immunohistochemistry

Immunostaining of normal adult human tissue samples with GPRA A variant specific antibody showed strongest expression in smooth muscle cells (SMC), such as SMC layer in bronchial and arterial walls in human lung and colon shown in Fig. 10A and E. In alveolar wall and alveolar macrophages of lung (Fig. 10C) intense staining of *GPRA* A was detected. In colon epithelium (Fig. 10D), strong basal staining was observed, whereas in the bronchial epithelium (Fig. 10A) and keratinocytes (Fig. 10F), only mild/weak immunostaining was detected.

Western blot analysis with GPRA A variant specific antibody revealed four intensive bands corresponding to molecular weights of approximately 50, 44, 42 and 40 kDa (Fig. 12A). In skeletal muscle only 50 kDa band was detectable whereas uterine muscle, colon epithelium and prostate showed similar expression patterns at 50, 44, and 40 kDa. The most intensive GPRA A expression was recorded in colon muscle with the major band at 42 kDa. Weak or nonexistent bands were found in spleen, kidney, and testis.

Immunostaining with GPRA B variant specific antibody revealed most intense expression of the protein in the epithelium of several mucosal tissues including bronchus, small intestine, and colon as shown in Figs. 11A, D, E. Contrary to the A variant, the expression of the B variant was strongest in the apical surface of the epithelium (Figs. 11A, D, E, and F). Alveolar walls and alveolar macrophages (Fig. 11C) showed strong and SMCs in several tissues mild immunostaining of the variant B.

With GPRA B variant specific antibody, two sharp polypeptide bands corresponding to molecular weights of approximately 39 and 25 kDa were revealed by Western blot analysis with the strongest expression in kidney (Fig. 12B). The specificity of the antibody was confirmed by the blocking experiment with 10x molar excess of the peptide used in immunization.

Cos-1 cells were transiently transfected with pCMV-myc GPRA A, B, Bshort, C, D, E, or F in order to study translocation of different GPRA variants *in vitro*. After 48 hours, cells were fixed and studied by cell based ELISA assay using myc antibody. 71% of recombinant GPRA

A and 52% of recombinant GPRA B were translocated to the plasma membrane while all the other variants were located in cytoplasm (Table 15).

3. Characterization Of GPRA as a Genetic Regulator of Asthma Related Traits

A. Detection of GPRA Polymorphisms Associated to Asthma Related Traits

To detect sequence variations in *GPRA* associated to asthma related traits, we sequenced its verified exons and exon-intron boundaries in four patients homozygous for the haplotype significantly associated with high serum total IgE level among the Finnish asthma families. Comparison between patients and one control individual from the same data set were made to the reference sequence (genomic contig NT_000380) in the public data base.

A total of 14 SNPs were found distributed in seven exons of *GPRA* (Table 7). Four of the SNPs were predicted to cause a non-conservative amino acid change: Asn>Ile located in the first extracellular loop, Ser>Arg in the third cytoloop, Gln>Arg and Thr>Ile in the C termini, one in each alternative splice variant. Six of the SNPs were located in the 3' UTRs of the A, B, and C variants of the gene.

B. Association Analysis

We have previously defined a susceptibility haplotype that significantly associates with asthma related phenotypes (best marker-wise P value based on 10,000 permutations 0.001) (Fig. 3, gray area). The region covers the genomic region between *GPRA* exons 2b and 5. We replicated the association analysis with five SNPs genotyped in the exons of *GPRA* in the same data set (304 disease associated and 220 control chromosomes). Best haplotype associations of *GPRA* for high serum IgE level when one gap was allowed in the haplotype patterns are shown in Table 8. The best observed associations reached χ^2 values of 5.7-10.3. Only one of markers (SNP591694) was part of the previously determined susceptibility haplotype and the same amino acid change, 107Ile, showed also significant association to high IgE level together with other *GPRA* SNPs.

To study whether the overall haplotype distribution in *GPRA* differs in high IgE associated compared to that in control chromosomes, we used a permutation test. HPM analysis was done using the following parameters: maximum pattern length 5 markers, one gap allowed for missing data and possible errors, and chi-square threshold for the association >5.0. The

observed scores (=number of qualified haplotype patterns spanning across the marker) for associated haplotypes varied from 6 to 14. Permutation test showed statistically biased haplotype distribution for high IgE. Based on 10,000 simulations, the best marker-wise P value ≤ 0.01 was observed for a silent SNP (in position 640,764) in exon 5. P ≤ 0.02 was observed for three markers in the middle of the haplotype and P < 0.04 for all the markers in the haplotype.

Example VII

This example describes a second gene occurring partially within the AST-1 locus and overlapping the GPRA gene.

A. Materials and Methods

1. Exon Prediction

Exon predictions were performed by GENSCAN (<http://genes.mit.edu/GENSCAN.html>) software using genomic clone NT_000380 (Polvi *et al.* 2002). Predictions were performed for both unmasked and masked sequence. In the latter, the repetitive sequences were masked using the Repeat Masker Service (<http://repeatmasker.genome.washington.edu/>).

2. Reverse Transcriptase PCR

The exon-specific primers were designed based on exon predictions (Table9, Figure13). First strand cDNAs were synthesized from 1 μ g of commercially available poly A⁺ RNAs from human lung, testis, kidney and fetal liver (BD Clontech) by the SMART RACE cDNA Amplification Kit (BD Clontech, Palo Alto, CA) according to manufacturer's instructions. Alternatively, Marathon-Ready cDNA (Clontech) or cDNA from Multiple Tissue cDNA Panels I and II (Clontech) were used as templates in some PCR amplifications.

All the PCR amplifications were performed in 25-50 μ l volumes using 2.0-5.0 μ l cDNA as template, 2 mM MgCl₂, 0.2 mM dNTPs (Finnzymes), 0.2 mM of each primer, 0.15-0.4 U AmpliTaqGold, and 1x PCR Gold buffer (Applied Biosciences) under the following conditions: 94°C for 5 min, 35-40 cycles of 94 °C for 30 s, 58 °C –62 °C for 40 s, 72 °C for 1 min 30 s followed by 72 °C for 10 min. PCR products were analyzed on 1 % agarose gel and extracted from the gel using QIAquick Gel Extraction Kit (Qiagen). Purified PCR

products were analyzed by automated sequencing with dye-terminator chemistry (Megabase 1000 sequencer, Molecular Dynamics).

3. Cell Culture and Isolation of Poly A⁺ RNA

In addition to commercially available poly A⁺ RNAs, RNA from Epstein Barr virus infected lymphoblast cell lines of the patients who were homozygous, heterozygous, or non-carriers of AST1 were cultured for expression studies. Lymphoblasts were cultured in RPMI 1640 medium (Gibco BRL) supplemented with 1 mM sodium pyruvate (Gibco BRL), 10% FCS (Biological Industries), and 1% Penicillin/Streptomycin (GibcoBRL). Poly A⁺ RNA was isolated by Dynabeads mRNA DIRECT Kit (Dyna) according to the manufacturer's instructions.

4. Northern Blot Hybridization

The AAA1 specific probe (mixture of the variants I, III, IV, VII, and X) was generated by RT-PCR using human lung, testis, kidney and fetal liver poly A⁺ RNAs as templates. The probe was radiolabelled with α [³²P]-dCTP using RediPrime Kit (Amersham Biosciences) according to the manufacturer's instructions. A Human Multiple Tissue 12-lane Northern blot, Fetal Multiple Tissue Northern blot, and Human Multiple Tissue Expression Array 2 (BD Clontech) were prehybridized in ExpressHyb solution (BD Clontech) for 1 h at 68 °C followed by hybridization with the specific probe for 2-5 h at 68 °C. Herring sperm DNA (100 µg/ml) was used as the blocking reagent. Filters were washed with 2 x SSC and 0.05 % SDS at room temperature and exposed to X-ray film at -20 °C for 1 - 5 days.

5. Genotyping

Genotyping of four SNPs that are located either in the coding region or near the exon-intron boundaries of AAA1 was done 1) using altered restriction sites (SNP_538567 or SNP_574953, Table 1B) or 2) single base pair extension (SBE) with the chemistry of Molecular Dynamics on a Megabase 1000 sequencer (Molecular Dynamics) according to the suggestion made by the manufacturer (Table 10). Allele calling was performed by using the MegaBACE SNP Profiler software (Molecular Dynamics).

All the markers were in Hardy-Weinberg equilibrium and observed Mendel errors were less than 0.1%.

6. In vitro translation of the AAA1 and characterization of AAA1 antibody

In order to investigate whether AAA1 is translated to any polypeptide, *in vitro* translation experiments were performed (Figure 21). Capped RNAs of AAA1 gene variants defined by SEQ ID NOS: 16 and 22 were transcribed from DNA constructs with the aid of T7 RNA polymerases (mMESSAGE mMACHINE system, Ambion, USA). Translation was performed with rabbit reticulocyte lysate translation machinery (Riboprobe *in vitro* translation system, Promega) in the presence of S³⁵-labelled methionine in the reaction mixture. The *Xenopus* elongation factor α (pTRI-Xef) DNA template was used as a positive control for transcription and translation. In negative control, water was used instead of DNA. The translated polypeptides were detected by autoradiography after Tris-Tricine SDS-PAGE.

B. Results

1. Characterization of AAA1 at RNA Level

A. Genomic Structure of Different Splice Variants of AAA1

The human AAA1 gene spans 520 kb of the genomic contig NT_000380 (nucleotides 163615–684776) and it is divided into 18 exons (Figure 13 and Table 11). The cDNA sequences of AAA1 splice variants were assembled by using RT-PCR with primers designed for predicted exons. The genomic location of the primers is shown in the Figure 13 and the sequence in the Table 9. Using cDNAs from human lung, kidney, testis and fetal liver as templates, twelve splice variants were identified repeatedly (variants I–XII in the Figure 14). All variants share exon 6, but use alternative exons to encode the 5' and 3' ends of the transcript. In six variants out of ten, the sequence flanking the putative ATG translation initiation site (GCCATGC) contains the –3 purine, but not the +4 guanine residue, of the Kozak consensus sequence.

B. Expression Profiling of AAA1 Using RT-PCR and Northern Blot Hybridization

Northern blot analysis of a multiple tissue expression array showed that AAA1 is expressed in several human tissues (Figure 16). Strong expression was seen in testis, brain, placenta, lung, heart, skeletal muscle, kidney, liver, fetal liver, and fetal lung. In multiple tissue northern blots, two main transcripts (2.4 and 7.5 kb in size) were detected (Figure 17). Additionally, all sample lanes showed extensive label streaming which can be an indication of several low-abundant transcripts with different sizes.

By RT-PCR exceptionally strong splicing and tissue specific differences in AAA1 expression could be found (Figure 18). For example, expression of splice variants I and IV was particularly strong in the lungs consistent with the association between AAA1 and asthma.

In the *in vitro* translation experiment, the *Xenopus* elongation factor α (pTRI-Xef) DNA template used as a positive control resulted in the synthesis of a 50-kDa polypeptide as expected. However, neither of the two investigated constructs was translated into polypeptides strongly arguing that AAA1 functions as a non-coding RNA gene (Figure 21).

To further investigate the translation of the AAA1 gene, a polyclonal antibody against the constant region (YVRRNAGRQFSHC) of the gene product was produced in rabbits. AAA1 peptide synthesis and antibody production were purchased from Sigma-Genosys Ltd (London Road, Pampisford, Cambridge). To test the specificity of the antibodies, Glutathione S-transferase (GST)-fusion proteins for AAA1 were produced with the pGEX 4T-3 GST fusion expression vector (Amersham Biosciences) according to the manufacturer's instructions. (Figure 21). The antibody displays high affinity against the recombinant AAA1 protein produced in bacterial lysate with no cross-reactivity between the GST construct alone. In spite of that, the antibody did not reveal any reactivity either in Western blots (spleen, skeletal muscle, uterine muscle, colon muscle, colon epithelium, kidney, testis and prostate) or in immunohistochemistry (bronchial tissue, HepG2 cell line) (data not shown).

In addition, formalin fixed, paraffin-embedded specimens of normal adult human bronchus tissue and HepG2-cells (positive for AAA1 RT-PCR) were used for immunohistochemistry. Immunohistochemical analyses were performed using rabbit AAA1 antibody and the ABC method (Vectastain Elite ABC kit, Vector Laboratories, Burlingame, California). Finally, AAA1 does not appear to have a mouse counterpart. Concluding from our data, AAA1 may be a non-coding RNA gene and is unlikely to encode a functional protein.

C. The Predicted Structure of the AAA1 Peptides

The cDNA sequences of twelve AAA1 splice variants are shown in Fig. 20. All predicted AAA1 proteins are small peptides (size from 34 aa to 74 aa) that do not show significant

identity to any known modular structures or motifs. All isoforms contain the same core sequence (AYVRRNAGRQFSHCNLHAHQFLVRRKQ) flanked by alternative amino- and carboxyterminal tails (Figure 15).

2. Characterization Of AAA1 as a Genetic Regulator of Asthma Related Traits

A. Association Analysis

AST1 covered the exons 3 to 10 of AAA1 (Figure 13). For the association analysis we chose one exonic SNP (SNP 517278) and four additional SNPs near exon-intron boundaries of AAA1 (Table 12). All the polymorphisms were identified previously (Table 3). The analysis for high total serum IgE level was done in the same data set as previously (304 disease associated and 220 control chromosomes, Table 2 and 8). Best haplotype associations of AAA1 are shown in Table 12. The best observed associations reached the χ^2 values of 8.9-13.6 and the permutation test showed statistically biased haplotype distribution for high IgE. Based on 10,000 simulations, the best marker-wise P value ≤ 0.0001 was observed for SNP_517278 in exon 10. Corrected P value for testing multiple markers simultaneously was 0.0002.

B. Variable alternative splicing for AAA1 depending on genotype

To study further whether AST1 effects on the expression of AAA1 we studied lymphoblast cell lines from the asthma patients who were homozygous, heterozygous, or non-carriers of AST1. Using RT-PCR with the primer pair SCF10 and ASKAR (Table 9) significant differences were found between patients and the differences were dependent on genotype (Figure 19). Only the non-carrier of AST1 processes normal amount of the exon 6-10b transcript, whereas the homozygote and heterozygotes show either an absent transcript or smaller splice variants. Differences in expression patterns suggest that AST1 can effect splicing of the gene and thereby, increase the risk of asthma related diseases among AST1 carriers compared to that among AST1 non-carriers.

REFERENCES

- Dantzker, D.R. *et al.* 1987. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *Am. Rev. Respir. Dis.* **136**: 225-243.
- Becker, K.G. *et al.* 1998. Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 9979-84.
- Daniels, S.E. *et al.* 1996. A genome-wide search for quantitative trait loci underlying asthma. *Nature* **383**: 247-50.
- Dizier, M.H. *et al.* 2000. Genome screen for asthma and related phenotypes in the French EGEA study. *Am. J. Respir. Crit. Care Med.* **162**: 1812-8.
- Jacob, H.J. *et al.* 1992. Genetic dissection of autoimmune type I diabetes in the BB rat. *Nat. Genet.* **2**: 56-60.
- Laitinen, T., *et al.* 2001. A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat. Genet.* **28**: 87-91.
- Leaves, N.I. *et al.* 2002. A detailed genetic map of the chromosome 7 bronchial hyper-responsiveness locus. *Eur. J. Hum. Genet.* **10**: 177-82.
- Malerba, G. *et al.* 2000. Linkage studies to asthma and atopy phenotypes on chromosomes 7, 12, and 19 in the Italian population. *Am. J. Hum. Genet.* **67**: 330.
- Mathias, R.A. *et al.* 2001. Genome-wide linkage analyses of total serum IgE using variance components analysis in asthmatic families. *Genet. Epidemiol.* **20**: 340-55.
- Ober, C. *et al.* 2000. A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am. J. Hum. Genet.* **67**: 1154-62.
- Polvi A *et al.* 2002. Physical map of an asthma susceptibility locus in 7p15-p14 and an association study of TCRG. *Eur. J. Hum. Genet.* **10**, 658-665 (2002).

Remmers, E.F. *et al.* 1996. A genome scan localizes five non-MHC loci controlling collagen-induced arthritis in rats. *Nat. Genet.* 14: 82-5.

Satsangi, J. *et al.* 1996. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat. Genet.* 14: 199-202.

Sawcer, S., *et al.* 1996. A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat. Genet.* 13: 464-8.

Sevon, P. *et al.* 2001a. TreeDT: Gene mapping by tree disequilibrium test. In KDD-2001. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 365-370. Editors: Provost, F. and Srikant, R. Publisher ACM press, San Francisco, CA USA
(www.acm.org/sigkdd/kdd2001)

Sevon, P. *et al.* 2001b. TreeDT: Gene mapping by tree disequilibrium test (extended version). In the publication series Report C-2001, Nr 32:6-7 published by Department of Computer Science, University of Helsinki

Toivonen, H.T. *et al.* 2000. Data mining applied to linkage disequilibrium mapping. *Am. J. Hum. Genet.* 67: 133-45.

Wjst, M., *et al.* 1999. A genome-wide search for linkage to asthma. German Asthma Genetics Group. *Genomics* 58: 1-8.

Xu, J. *et al.* 2001. Genomewide Screen and Identification of Gene-Gene Interactions for Asthma-Susceptibility Loci in Three U.S. Populations: Collaborative Study on the Genetics of Asthma. *Am. J. Hum. Genet.* 68: 1437-1446.

Xu J *et al.* (2000). Major genes regulating total serum immunoglobulin E levels in families with asthma. *Am J Hum Genet.* 67:1163-73.

- Yokouchi, Y. *et al.* 2000. Significant evidence for linkage of mite-sensitive childhood asthma to chromosome 5q31-q33 near the interleukin 12 B locus by a genome-wide search in Japanese families. *Genomics* **66**: 152-60.
- Daly, M.J. *et al.* 2001. High-resolution haplotype structure in the human genome. *Nat Genet.* **29**: 229-32.
- Dantzker, D.R. *et al.* 1987. Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease (COPD) and asthma. *Am. Rev. Respir. Dis.* **136**: 225-243.
- Johnson, E.N. *et al.* 2002. Heterotrimeric G protein signaling: role in asthma and allergic inflammation. *J Allergy Clin Immunol.* **109**: 592-602.
- Kauppi, P. *et al.* 1998. Verification of self-reported asthma and allergy in subjects and in their family members volunteering for gene mapping studies *Resp. Med.* **92**: 1281-1288.
- Laitinen, T. *et al.* 1997. Genetic control of serum IgE levels and asthma: Linkage and linkage disequilibrium studies in an isolated population. *Hum. Molec. Genetics* **6**: 2069-2076.
- Laitinen, T. *et al.* 2001. A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat. Genet.* **28**: 87-91.
- Michel U 2002. Non-coding ribonucleic acids--a class of their own? *Int. Rev. Cytol.* **218**: 143-219.
- Numata *et al.* 2003. Identification of Putative Noncoding RNAs Among the RIKEN Mouse Full-Length cDNA Collection. *Genome Res.* **13**: 1301-06.
- Palczewski, K. *et al.* 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**: 739-45.
- Park, Y. *et al.* 2002. Identification of G protein-coupled receptors for *Drosophila* PRXamide peptides, CCAP, corazonin, and AKH supports a theory of ligand-receptor coevolution. *PNAS* **88**: 11423-28.

Rana, B.K. *et al.* 2001. Genetic variations and polymorphisms of G protein-coupled receptors: functional and therapeutic implications. *Annu. Rev. Pharmacol. Toxicol.* 41: 593-624.

Although the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. The above examples are provided to illustrate the invention, but not to limit its scope; other variants of the invention will be readily apparent to those of ordinary skill in the art and are encompassed by the claims of the invention. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents. Any embodiment, feature, step or element described above can be used in combination with any other unless otherwise apparent from the context. All publications, references, and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

Table 1. Primers that were used in SNP genotyping with SBE method (A); primers, restriction enzymes, and the length of the digested PCR product of corresponding alleles used in SNP genotyping with altered digestion (B); primer pairs and the length of the corresponding PCR product used in genotyping of the small insertions and deletions (C).

Table 1A.

SNP ID	Nucleotide	Type ¹	Fw primer	Rev primer	PE PRIMER
NT 000380.3					
dbSNP: 323926	506401	C/T	CCCCCTGCTCTCTCTCTT	GGGATTATCCATTTCCTCCA	ATGCCAGCATCATCTTCCTG
dbSNP: 323906	509240	C/G	GATCTTCCATTTCACCAAGCA	GATGGATATTTAGTCCCCCTTA	CACACTAGCAAAAGTCAGGCAAT
dbSNP: 323917	515224	C/G	TGAGTTGATGAATCTGCTGGA	GGCAATGCCCAATTAAGAGGA	CACGAGCCCTGGATTACCTAC ²
dbSNP: 143266	515632	A/G	CAATGGATCACTGCAGCCTA	TTCCACCCATTAGTTTTC	AGAAACAAGCTATGCCAGAGACTG ²
dbSNP: 324377	529556	A/C	AACACAGCCCTCCAGACACT	ATACCTGGCCAGCGGTTTA	TGGAAAAAATTTGAACATTATTGGG
dbSNP: 324373	531632	A/G	CTATTGGGCCCTGGAGACTT	ATGCAGCTTGCAATTGTTAC	GGTCTCTGGGACCCCTTC ²
dbSNP: 182718	543580	C/T	AGGGATCATGTCTCCAGCAC	TTAAATGCCCACTAGCACCA	GAGCAGAGATCAGAAACAGGCT ²
dbSNP: 324384	555608	A/G	TGTGTCTCTCTTTGTGTTTCATA	TGAAAGCAGGGGTACCATTT	GCAGGAA CAGAAAACGAAATACC ²
dbSNP: 324396	563704	C/T	CACAGGGAGGAAAGGAACAG	GATGGGATTGAGGAGCTTGA	ACAGTGGATGGTTTCTTGGC
dbSNP: 324981	591694	A/T	GGCCATCTGTATAAAGCAGGA	TGATGCATAGGAATGCAAGG	AGTCTOCAGTGAATCGCCAA ²
dbSNP: 325456	617392	C/T	GATGAAAAGGGAACCATTCCT	TGCTTTTCTGTGTCCATTG	ATGCCCTAAATTCCTAAAACCAATGA
dbSNP: 62911	638799	C/T	TGACATGTTCCCTGCATTTG	AGAGGGCTTTGTGCTCTCTG	CACCCCTCCCTGGTGACCT ²

¹ Polymorphism detected in assays employing primers specified in this table

² extension primers (Pe) designed in reverse orientation

Table 1B. Primer, restriction enzymes, and sizes of the corresponding alleles after digestion used in SNP genotyping of AST1.									
Marker	Forward primer	Reverse primer	Enzyme	Product size*	Allele 1	Product size*	Allele 2	Note	
SNP_490331	gcctgccttagagacatcag	tggtgcctcttcattcttg	HpyCH4V	180+110	c	180+90+20	a	special combinations of fragments when heterozygous for the SNPs 490331 or 490390	
SNP_490390	gcctgccttagagacatcag	tggtgcctcttcattcttg	HpyCH4V	180+110	a	180+80+30	g		
SNP_490820	ccaattctatccaaagc	taaacatggctgcacattgctcactggatggc	DraIII	330	t	250+80	g	plasmid tail with the DraIII restriction site	
SNP_513659	aactgactaaactagggtgccacgtcg	tgattgcatcttgctctaagtg	MnlI	220	g	170+50	c	plasmid tail	
SNP_514743	ggaaactgggaagtcacaccca	tctccctcagaaaagccaga	SspI	420	g	210+210	a		
SNP_516094	ccctgcaataaactctcataatfaaa	gggtggattacagggcatgag	FauI	400	t	210+180	c		
SNP_516174	gaagtggaacttactgcatcaaga	gggtttcacctggttagc	SfiI	400	t	200+200	g		
SNP_516287	gctctagaacacccgtagcat	gcctcagctctggagtagc	BsrBI	380	a	180+200	g		
SNP_520598	gcaaaagttgaggggagtc	accatgcctgcctcatagtaa	NlaII	300	c	130+170	t		
SNP_521640	aggaaacctccacagcag	cccaaaagccttgagaaaaat	BipI	360	a	180+180	g		
SNP_522363	tgcaacagcttttgaggga	tgtaacatccccaggagct	EcoNI	380	g	220+160	a		
SNP_526991	ctttaggggaagcgctgg	aagccactaatatttggagtg	BsaWI	400	c	230+170	a		
SNP_528709	tgctctttttgacagccaatc	ggctgctttgatctgagpac	NcoI	360	t	180+180	g		
SNP_529142	ccataagccctgcctatttt	gtgggcactgaggcacagat	TaqI	410	a	240+170	g		
SNP_529820	caaatcctgtctctctgct	tacccagagcctgagaaagg	BsaAI	400	t	220+180	c		
SNP_530177	cccttcacgctcctggga	gtgagcgtgagctgcgctat	StyI	380	t	230+150	c		
SNP_538557	ctcagaacactgctgccaga	gctctgggatgcgttaactt	EcoNI	380	g	220+160	a		
SNP_541906	ttgatcggccaaalggaac	ctaaacaaaaggggctgcaa	AflII	350	c	180+170	a		
SNP_546333	gcattgaaatcacagcacttg	ctcccgagttcaagcaatc	HaeIII	290+90	a	190+100+90	g		
SNP_563585	ccctacaagttacccctgcaca	cagaggttggatggagtg	HindIII	375	c	180+185	t		
SNP_563930	tcaagctctcaatcccatc	gtggagaagtgctcttcca	SitI	330+40	g	150+180+40	a		
SNP_564782	aaaggcagccagtgctgtt	ggcagagatgccacagtga	MspI	400	a	170+230	g		
SNP_574953	ttctagcctgctacccttla	aagcacaagcatcaggltcca	AhdI	365	a	180+185	g		
SNP_585883	tcaacagagccalgattaga	gcagatggaggtagagctggtt	SfiI	370	g	180+190	c		
SNP_640764	aactgactaaactagggtgccacgtcg	tagcacaatgctgcctcat	SapI	340	t	290+50	c	plasmid tail	
SNP_647327	ttccatacatgccatgctctt	gcctgagaagaacccaaagg	BstEII	455	c	210+245	t		
SNP_662764	aggcaagtcacctgctcttc	gtctgtgcatctccaggt	MboI	455	a	240+215	g		
SNP_662803	catggagaagggaaggtcagg	ctagcactggcactgcccta	DdeI	270	c	220+50	t		
SNP_663133	atgactgcatgactgctta	tcctgaacaccagggtatct	SnaI	400	t	140+260	c		

Table 2. HPM permutation test done using the following parameters: maximum pattern length 40 markers, one gap allowed for missing data and possible errors, and chi-square threshold for association 13.0. The observed scores (=number of qualified haplotype patterns spanning across the marker) and marker wise P values based on 10 000 permutations are shown. The critical region (=AST1) is highlighted.

Marker	HPM Score	Marker wise P value
NM4	0	1
NM5	0	1
D7S683	0	1
D7S656_2	0	1
NM2	0	1
MIT_MH26	0	1
NM8	0	1
NM9	0	1
NM11	0	1
G42099	0	1
SNP_490331	0	1
SNP490391	0	1
SNP_490820	0	1
SNP_506401	0	1
SNP_509240	0	1
NM50	0	1
SNP_513659	0	1
SNP_514743	0	1
SNP_515224	0	1
SNP_515632	0	1
SNP_516094	0	1
SNP_516174	0	1
SNP_516267	0	1
NM49_516568	0	1
NM51_517022	2	0.0139
G42102_518794	7	0.0105
NM53_519564	12	0.0091

Table 2. Continued

Marker	HPM Score	Marker wise P value
SNP_520598	15	0.0083
SNP_521640	20	0.0079
SNP_522363	23	0.0075
SNP_526991	23	0.0076
SNP_528709	30	0.007
SNP_529142	33	0.0066
SNP_529556	40	0.0062
SNP_529820	33	0.0061
SNP_530177	34	0.0059
SNP_531632	34	0.0054
D7S497_538277	40	0.0052
SNP_538567	38	0.0052
SNP_541906	36	0.0054
SNP_543580	35	0.0054
SNP_546333	35	0.0054
SNP_555608	35	0.0051
NM45_560804	35	0.0049
SNP_563585	35	0.0049
SNP_563704	28	0.0046
SNP_563930	23	0.005
SNP_564782	16	0.0063
SNP_574953	10	0.0071
SNP_585883	4	0.0093
SNP_591694	3	0.0098
SNP_617392	3	0.0085
SNP_638799	0	1
SNP_640764	0	1
G42097	0	1
SNP_647327	0	1
SNP_662764	0	1
SNP_662803	0	1
SNP_663133	0	1
NM13	0	1
NM46	0	1
D7S484	0	1

Table 1. Polymorphisms and flanking sequence

Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Polymorphism and flanking sequence	H1	H2	H3	H4	H5	H6	H7	Location in <i>GPRA</i> and/AAA1 exons	Amino acid change
SNP	1	509783	ACAAATACATAGTAACTTAG(C>T)GGCAATTAATGTAGC	C	C	C	C	C	C	T		
SNP	93	509875	TTGAGCATGAAAAAGAGC(A>G)AAATACAAAAAGA	A	A	A	A	A	A	G		
Insertion (NM50)	184-185	509866-67	GATAGAACCAACTTAAAGATAGTAGTGGCAATATTTCCCAACTCT	no ins	no ins	no ins	ins	ins	ins	ins		
SNP	918	510700	TCCCTCGCCAAATACAAAATTCGAG(G>A)ATGCTTAAGTCCCTTACATAAA	G	G	G	G	G	G	A		
SNP	983	510765	CTACATACACCCCTCTGTAT(A>T)CTTTAAGTCACTCTCTAAATACCT	A	A	A	A	A	A	T		
SNP	987	510769	CTACATACACCCCTCTGTAT(A>T)CTTTAAGTCACTCTCTAAATACCT	A	A	A	A	A	A	T		
deletion	1202-1206	510984-88	ATATTATATAAATCAATCTTAAGTAAAGACAGGAGAGAAAAATATA	no del	no del	no del	no del	no del	del	no del		
SNP	1542	511324	TATCAAAAGTTTATAAATTCGTGTGAT(C>T)TTGGAATCACTGTAGCTTTATC	C	C	C	C	C	C	T		
SNP	1710	511492	CACCAATATATATATGATTTTGT(G>A)GTACTTTCTTATGTTTAT	A	A	A	A	A	A	G		
SNP	1818	511600	CAATGGCTTTAGGAATCAAGCAGG(C>T)ATGAATGTAACATGTAAA	C	C	C	C	C	C	T		
SNP	1927	511709	AAGACATTAATCAATTTTGT(C>A)TACACTGTGCTGTAGTCAACAAT	A	A	A	A	A	A	T		
SNP	2254	512036	ATTTCTTCTCTAATGTGTTT(C>A)TATGTTTGTGTTTATGATAAT	T	T	T	T	T	T	C		
SNP	2837	512719	CAACACAGAAATTAATTCCT(C>G)ACAGTCTGTAGSCTGAAGTTT	G	G	A	G	A/G	A	G		
SNP	3877	513659	TCACACAGGATTAAGCCTG(C>A)AGTCTTTTTCAGAGAAGA	G	C	C	C	A/C	C	C		
SNP	4012	513784	TTACACCTTCAAGATGGCTATTC(C>A)CTATAAATAATATACTAA	C	A	C	C	A/C	C	C		
SNP	4631	514471	CAACAGCAATTCAGTACAGAGACTTTT(C>G)TCGGAATCAGCTGCCATAAA	T	T	C	C	C	C	C		
SNP	4689	514471	CAAAAGAAATTCAGTACAGAGACTTTT(C>G)TCGGAATCAGCTGCCATAAA	T	T	C	C	C	C	C		
SNP	4961	514743	CCGTGCGACCCCATGACATAATTA(G>T)TAAATGAATAT	C	C	C	C	C	C	G		
SNP	5442	515224	TATCTTATGAAAGTCCGGG(C>G)GTAGGTAATCCAGGGCTCGTCTG	A	G	A	G	G	A	A		
SNP	5634	515416	CAGGAAGGAAGAGGGAGAA(G>A)GACAGAGGGGCTCATATAGCTCTTG	C	C	C	C	C	G	G		
SNP	5850	515632	AGAAATAGAAATTAAGGAA(G>G)CAGTCTCTGGCATAG	A	G	A	G	G	A	A		
SNP	6312	516084	AATGTCGGGAGCAATGCTCTAGAA(C>T)CCGCTGAGATAGAC	C	T	C	T	T	C	C		
SNP	6392	516174	ACTGCGAGTACTGAC(G>T)AATCATCACTGGTTTTT	G	T	G	T	T	G	G		
SNP	6485	516287	GAATTTGTTATTTCAAAAG(G>A)AGCGCCAGCGCGCGTGG	G	A	G	A	A	G	G		
SNP	6522	516304	TGGCTCATCTGTAATCCG(A>G)CAGCTTTGGAGGCGCAAGCGGG	C	G	C	G	G	C	C		
SNP	6646	516428	GGCGTGTGGGCGGCGG(A>G)GTGGAGCGGCTGTAGTCCAGC	A	G	A	G	G	A	G		
SNP	6739	516521	CGGAGCTCGAGTGGAGCGAGATCG(G>A)CCACTCCCTCCAGCTGGG	G	A	G	G	G	G	G		
SNP	6760	516542	ACTGCCCTCAGCTGGGTTT(C>G)GACAGAGCAAGCTCC	T	C	T	C	C	T	T		
SNP	6786-6817	516568-69	AGCAAGACTCCGCTCAAA(TAAAB>TAAAT)GCAATTTTTTTTCACTTTAG	8	7	8	7	7	8	8		
lelellon	6821	516603	ATAAATAAATAAAGCA(TTTT)TTTTTTTCACTTTAGCCAGC	no del	del	no del	del	del	no del	no del		
SNP	7125	516907	TGAAGCCATTTTGG(C>A)CATCTTGTAACTTCTTAGAA	C	C	C	C	A	C	C		
SNP	7229	517011	CCAATTAAATTTATCTTAC(C>T)CATCTTAAAGTCTTGTGTTT	C	T	C	T	T	C	T		
SNP	7240-7243	517022-25	ATGCAATAAATAATG(C>G)GGAATTAAGAGTAA	no del	del	no del	del	del	no del	no del		
SNP	7277	517059	ATGCAATAAATAATG(C>G)GGAATTAAGAGTAA	C	C	C	C	G	C	C		
SNP	7303	517085	AAGAGTAATCCCGATG(C>T)AGTGTTCACAAAGAAAGAAAGAA	G	T	G	T	T	G	C	AAA1 exon 10	no change (3'UTR)
SNP	7305	517087	AAGAGTAATCCCGATG(C>T)AGTGTTCACAAAGAAAGAAAGAA	G	T	G	T	T	G	C	AAA1 exon 10	no change (3'UTR)
lelellon (NM52)	7306-7308	517088-90	AAGAGTAATCCCGATG(C>T)AGTGTTCACAAAGAAAGAAAGAA	no del	del	no del	del	del	no del	no del	AAA1 exon 10	no change (3'UTR)
lelellon(NM52'241)	7334-7335	517116-17	CAACAAGAAAGAAAGAAAGAAATATTTTGGCGCAGATTCAAATC	no del	no del	no del	del	del	no del	no del	AAA1 exon 10	no change (3'UTR)
SNP	7486	517278	ACCCTCTATACATACCTGAA(C>T)GACAGAACTTGAAGCT	C	T	C	T	T	C	C	AAA1 exon 10	no change (3'UTR)
SNP	7550	517332	AGGTAAGCCAGAGTGTGAGG(C>A)TCCAGGTCGCTGTGGAGTT	G	A	G	G	G	G	G	AAA1 exon 10	no change (3'UTR)
SNP	8490	518272	CTCTATATGCTCTTTTATATA(C>T)ATATTTTTCCTGTGTGAGTT	C	T	C	T	T	C	C		
SNP	9012-9035	518794-817	TGTTGTGTATGTATGTTGAT(C>T)ATATTTTTCCTGTGTGAGTT	12	10	12	10	10	12	12		
lelellon	9199-9201	518981-83	TTTGTGAGTGTGTTGCAATTTATCTTCTCTCACTGTGACTTACTGAA	no del	del	no del	no del	no del	no del	no del		
lelellon	9355-9358	519137-38	TTGGTAGTATTTTTTTTTCAGAAAGTCTTGAAGTAA	no ins	ins	no ins	ins	ins	no ins	no ins		
SNP	9649	519431	TCTATTTTTGCTGTGTTATGTTT(C>G)GAGTCTCTGAGTGCACATA	T	G	T	G	G	T	T		
SNP	9782-9785	519564-67	CTTCAGAACTCTAAGTCTGTGTTATGTTATGTTATTTTGGCTTT	no del	del	no del	del	del	no del	no del		
SNP	10816	520598	AGAAAAGCTGCTTCTGCCCCAT(C>G)GAGAGGAAGTATGAGAAATAGA	T	C	A	C	C	T	C		
SNP	11858	521640	TTCCCAAAACACATGGGCTGA(A>G)CAGGTAAAGTGGATCTT	A	G	A	G	C	A	A		
SNP	12581	522363	GAATTTGGCGGCTCATTA(C>G)TTTCTCAGAGTCCCTGGGGG	G	C	G	C	C	G	C		
SNP	12845	526627	TCTAGAATTCCTCTTGTAGGCCATCAGT(C>A)ATCTCACTGAAATAGAC	T	T	C	T	T	T	T		
SNP	16893	526675	TATAAATAATAAT(C>A)ATATGATATTTTAAATAT	T	T	T	T	T	T	T		
SNP	16893	526762	CAGAAATGAAGGCGCAATCTTATGTT(C>G)GAGAGGCTGTGGAGCT	T	T	T	T	T	T	T		
SNP	17147	526929	TGGCAAGCTCTGCATAGTCC(C>A)TGAAGTGGAGATGGGAA	G	A	G	G	G	G	G		

Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Polymorphism and flanking sequence	H1	H2	H3	H4	H5	H6	H7	Location in GPRA and AAA1 exons	Amino acid change
SNP	17209	528991	TAGGCCATCCACGAGGCTGT<A>CJCGGGATCCCCTCGAGAATCTCAAGAAA	C	C	C	A	C	C	C		
SNP	17435	527217	GCGTCAAGAATTAAAGTGAATAC(A)>GJTGAGGTGCCTCATTTTC	A	G	A	A	A	A	A		
SNP	18363	528165	AGAACACACCACCATCCACGCTG(G)>ATATACTCACAGTAATGATGTG	A	G	G	G	G	A	A		
SNP	18927	528709	CATGTTCAATATCCCATG(G)>TATGGGGTATGATGTG	G	G	G	T	T	G	G		
SNP	18978	528760	TGAATGTGAGTTGAAGGACT(A)>GJTGTCACTCCACAGCTAGAGCTGTGAAA	A	A	A	A	A	A	A		
SNP	19268	529050	GACGACCTGGGTG(G)>CTGAATAGCACTCGGAAC	A	G	G	C	C/G	G	G		
SNP	19272	529054	CATAGGACACGACCTGGGTG(TA)>TTAGCACTGGGAAGCTGAAC	A	T	A	A	A	A	A		
SNP	19360	529142	ACTAGTGAGAAGAAGT(G)>AJAGGGAAGATAAAGTCAGAA	G	A	G	A	A	G	A		
SNP	19452	529234	CTTTGGGGAAGAGTGAAGAGGGT(G)>AJGCCCTCAGAGCTGAGTCCA	G	A	G	G	G	G	G		
SNP	19671	529453	CAGTCCCATTCATCCAAAAGAG(G)>AJAATGSACACATATAATCTA	G	G	G	G	G	G	G		
SNP	19712	529494	AGCCATAAATGATGG(C)>AJAATAACACAGCCCTCCAGA	C	C	C	A	A	C	C		
SNP	19774	529556	TTGAACAATATGGGG(C)>AJAATCACTGCCTCTCAATGAA	C	C	T	C	C	T	C		
SNP	20038	529820	CTTAGAACCATGTCTGTGATAT>CJGTGAGTGCCTCAATGAA	T	A	T	T	T	T	A		
SNP	20089	529871	TTGAATATATGATATACCCAGGT>AJTTTTAAACCGCTGCCAGGTATCC	T	A	G	G	G	T	A		
SNP	20309	530091	CTGCTTTGCACAAGACACAGATG(G)>ATCAGGGGCGAGTTACAGGAGTCTAC	G	A	T	G	G	G	G		
SNP	20395	530177	GCGAGCAGTGCCTGCT>CJTGGGAATGCTTTAATAACA	T	C	T	C	C	T	C		
SNP	20789	530571	CTCCGATCTCAGGTATCGCTG(G)>TICCTCAGCCCTCCCAAAGTCTGGGATT	G	T	C	G	G	T	G		
SNP	21850	531632	TTTGACTTTGCAGATCCCQ(C)>TJGAAGSGGTCCACAGAACCC	C	T	C	T	T	T	T		
deletion (NM54)	22122-22123	531804-5	TTTTCGTGCACCTCTGTGTGTGATGTGGAACCC	no del	no del	no del	no del	del/no del	no del	no del		
SNP	22475	532275	TGTCTGCTGGCTGGTGAAGATGTGCTGCAGCTGGCAGCAGC	T	T	T	T	T	T	C	AAA1 exon 9	no change (noncoding)
SNP	22493	532275	GTGAGAAGGTGTCACAGCTTGGCAGG(C)>GJGTGAGAGAAGTGGGGGA	A	A	A	A	A	A	G	AAA1 exon 9	no change (noncoding)
SNP	22715	532497	GTGCATGATATGTGGAC(G)>ATGCACCTCTCAACAGAGGGA	G	G	G	G	G	A	A	AAA1 exon 9	no change (noncoding)
SNP	22869	532651	AGCTAGATTGTGAGAGCTTTAGTT>CJGAAGGAGCAGGTGACC	A	T	C	T	T	T	T	AAA1 exon 9	no change (noncoding)
SNP	22934	532716	AAAAATGCAATACACACAGCTCCCTAG(A)>TJAATGATGACATGCTGTCTA	A	A	A	A	A	T	T	AAA1 exon 9	no change (noncoding)
SNP	24007	533789	TGTGTGATGTGGGGTTT>CJGGTGTGTGTGTTGGGGAT	G	G	T	C	G	T	T		
SNP	24264	534046	TCACATATCTCTTATCATCAAGT(G)>TJGAAGAAGAAATCTGTAATACTAT	G	G	T	C	T	T	T		
SNP	24869	534651	GGGCTCAACATTTTCAAGCACAGT>CJAATTGACATCAACATGGA	T	C	C	C	T	T	T		
SNP	26188	535980	CTCAGATCAATACAGATTC(C)>TJGTGTGTCAGCATCAACATGGA	C	C	C	C	C	C	C		
SNP	26356	536138	GTGAGGAAGGCCATTTAGAC(C)>TJGAAGTGAAGTGTCCCATTA	C	C	C	C	C	T	T		
SNP	26675	536457	AGCAAGTACTATTTCTGGGAAGATTC(A)>GJTGTGACCCCAATTAATTACA	A	11	10	11	10/11	A	G		
TAAA repeat (G42100)	26925-26988	536711-1150	CTAAATAGCA(ATAAA)10(TAAA)11TGATGAATGTAATATCT	10	11	10	11	10/11	A	G		
SNP	27404	537186	TTATCCATTTGGTTGAAC(C)>TJSAAACCCCATCTATCTCCA	C	C	C	T	T	T	C		
SNP	28187	537979	GTGATATAGAAAAGTGGCTTA(G)>AJCATAGTATCTAGTGTACAGTAGTGAOC	13	13	13	12	12	17	nd		
CA-repeat (D7S497)	28495-28564	538277-346	CTTTAAGTAACTCA(BAGACT)(CA)13-(CA)12TGA(CA)BCT(CA)BAGGCTCTC	13	13	13	12	12	17	nd		
SNP	28770	538552	AAITTAATGTGCCATCAACTGT(C)>TJCCAAATCCTCTACCG	C	T	C	C	C	T	T		
SNP	28785	538567	CCCAAAATCCTCTACCG(G)>AJGATAAGATCAAGCTACTCT	G	T	T	T	T	G	G		
SNP	28858	538640	TTGACATCTCAATATGAGAT>CJACTGAAGGACAGACACTAGTGAATC	T	G	T	T	T	T	C		
SNP	28866	538648	TTGAGACATCTCAATTTGAGGACTAGAG(G)>CJACAGACACTAGGTAATC	G	G	G	G	G	G	C		
SNP	31224	541006	TTGGGAGCCCTCAACACACTAAAGTG(G)>AJCCCTTTCACACTCAACTT	G	G	G	G	G	G	C		
SNP	31910	541692	AGCTCAGAAGCTCCAGGTAGACGA(G)>AJMCGCTCTTGCCTTAATCTTTAAAAA	G	A	G	G	G	A	A		
SNP	32124	541808	AATAACCCCTCTGATGGGCTTTC>AJAGACCCCATGATAGCTATA	C	C	T	C	A/C	C	C		
SNP	32185	541987	TTCTCACAAGAACTGATGGGCATGAAGAT>CJTGCAGTGACTGAAGTGTCTC	T	T	T	T	T	T	C	AAA1 exon 6	Asp>Ser
SNP	32978	542758	AAITTAATTTTGGAAATGAT>CJATGATTAAGATAAATAAAGAGATTGTTT	C	C	C	A	A	C	C		
SNP	33350	543132	CCACCCACCACTGCTGGCT(C)>AJAATGTTTATTTTATAGTACAA	G	G	C	A	A	G	C		
SNP	33798	543580	CTGAACAGTCTCCCTCA(G)>AJA GCCTGTTTCT GATCTCTGCT	G	G	G	C	C	C	C		
SNP	34362	544148	ACAAGTGAAGGCGGACGAGAGAAATATAG>CJTAAATAGGTTAAAGTTCCTCAAT	G	A	C	A	A	A	del		
deletion	34909	544681	AATGAAATTTTAAAGTACCCCTTJGGGTTTACATTTGCTTTTGGTACTT	no del	no del	no del	no del	no del	no del	del		
SNP	35559	545341	GTTTGAAGAAATCACTCACTGATA(C)>TJTAATAAGCTATCATGTAAACGGGCTC	C	C	C	C	C	C	C		
SNP	36551	546333	ATTTAAGATTCATGCA(G)>AJGCCAGGTGCAGTGGCTCA	G	G	G	A	G	G	G		
SNP	36909	546691	CAAGGCTTGCATAG(C)>ATATCCTGCAAAAATAAAGTTAGTT	C	A	C	A	A	A	A		
SNP	37327	547109	ATGAAAACCTGGGAGGGGG(A)>GJTGACACAGAGATGATATGA	G	G	G	G	G	G	T		
SNP	37415	547197	AGAGGAGTCAATAGTCTGG(G)>AJAGTGACACAAATCAGATTGGT	G	G	G	G	G	G	A		
SNP	37685	547467	GAGTGACATCTGGSGTATCA(A)>GJTTTTTTAAAGTCCCCAGAT	A	A	A	A	A	A	G		

Type of polymorphism	Position In SEQ ID NO:1	Position In NT_000380	Polymorphism and flanking sequence	H1	H2	H3	H4	H5	H6	H7	Location in GPRA and AAA1 exons	Amino acid change
SNP	37931	547713	AAAAAAACACATAGGTATAGTTCTC>TTTCAAACTTAACTTGGCCCTCTCCAG	C	T	C	C	C	T	T		
SNP	37959	547741	ATCTAACTGGCCCTCTCCAGGCGC>TTACGACACTGGGAGTGGCAATCAGC	C	C	C	C	C	T	T		
deletion	38850-38852	34	TTATCTCTGGAGCGCCCTTCCCTCCTCTGTTGTTGCTGACCTTAATAATG	no del	del	no del	no del	no del	no del	no del		
SNP	39314	549038	AACATTTGACAAATAATTTGTAATG>GATATAAAAAATCTCAACATGAATAA	G	A	G	G	G	T	A		
SNP	39343	549125	ATAAAGAAATCTCAACATGAATAAAAA>GTTACAGAAATAGAAAAATATATATG	G	G	G	G	G	T	A		
SNP	39927	549709	CTGCTTTTCTCTCTTTGATCCCAATCAGT>CTGGGCAACAAAGCTTGAACAAA	C	T	C	T	C	C	C		
SNP	45828	555608	AGCTCACACTTGTAACTGGGACATAGC>TTGGTATTTGCTTCTGTTCTCTGC	C	T	C	T	C	T	T		
SNP	50197	559979	GCTACTCGATCCAAAGATGGG>TTAAAGAGTGGAGGATATGC	G	T	G	G	G	T	T		
SNP	50334	560116	AGGTGGATACAGAGGTCAAGAGATCTC>GAAAGACCATCTCTGGCCCAACCTGG	G	G	G	G	G	A	A		
SNP	50493	560275	GTGGAGGTTGCACTGAGCCGAGACTGC>GATCACTGCACTCCACAGCTGG	G	A	G	G	G	A	A		
SNP	50632	560414	AGGATTTGGGGCATGAATAGAGAGCTG>ATCAGGTTGGAGACATCAACA	G	A	G	G	G	A	A		
SNP	50835	560617	CCTGAGCCTGTGGGAGGAGTACAGAG>GAGAGATGATCTTATCTGTGTC	G	A	C	G	G	C	G		
SNP	50955	560737	GCTGACAGCTTAAGTCAAGAGG>CTTGGATCAGAGATATCTTAATAGAT	G	C	G	A	G	G	G		
CA-repeat (NIMAS)	51022-51049	31	TTAACTTCAGTTTATAGACACAGC>GAAAGATCTTATTTATATACACTAG	wt	wt	wt	wt	wt	(CA)8	(CA)8		
SNP	51217	560989	GAACCTGAGTTTATAGACACAGC>GAAAGATCTTATTTATATACACTAG	G	A	G	A	A	A	A		
SNP	51476	561258	GCTCTGCCAGCTCACAGCA>GAAAGGCTCTTATTTATACAGAAAAAG	G	A	A	A	A	A	A		
SNP	51536	561343	CCAGTATGATAGCTGTATGTC>TTTTCATATAGGCATCAGTTAGACCC	C	C	C	C	C	C	T		
SNP	51661	561618	GACCAACAGAAAGAGAGCCTCAGGAC>CTGGGGCAGAGCATGAGAGCTAG	G	G	G	G	G	G	T		
SNP	51884	561666	GGCCACAGCATACAGCTAGGCG>TTGGTGGCAGCTGGGTTTAT	G	G	G	G	G	T	T		
SNP	51975	561757	TGCATGCCATACAAAGC>CTGGAAATCTCTACATCTGA	G	C	C	C	C	no ins	no ins		
insertion	52286-52287	69	TAGCATAGGCCATTCGATTCATCTCTCTCTCCGCCAGTACCTGATCTCTGA	no ins	no ins	no ins	no ins	ins	no ins	no ins		
SNP	52573	562355	TTCAATGTTCATAGCCACCAAAAGAGCTG>AGACCTGTGTGGCCACTCTCA	G	A	G	G	G	G	G		
SNP	52776	562558	TTTCAAGGATTAAGGGGAAGGCG>CTGGTAGAAAGGGGAAGTGGAGAGAGT	G	G	G	G	G	G	G		
SNP	53803	563585	GGTTCTTTTCGGAAGCTC>TTGCTGCTGGGTGTTCCGGA	C	C	C	C	T	C	C		
SNP	53922	563704	GTGATGGTTTCTTGGGTT>CTGTTGGGGCTCTCTGGC	T	C	C	C	C	A	C		
SNP	54148	563930	CAAGCATCTTGCAAGC>GATGGCTTCTCTGTGAGGGG	G	T	T	T	T	C	C		
SNP	54199	563981	ACACACAGTGGCCAGCTCCGCT>CTGGACCTGCTCAGTTAGGCTGTGTGG	T	T	T	T	T	C	C		
SNP	54641	564423	AGATCCCAAGTGGCAAGC>GATGAGATAGACCCG	G	T	T	T	T	C	C		
SNP	54751	564533	GCAGAACTTGGCAAGGCT>CTGAGGTGAGCAGCTGCA	T	T	T	T	T	C	C		
SNP	55000	564782	TTGTCTCTCTGTTCTCCCG>GATGAACTGGGCGCTG	G	G	G	G	G	A	G		
SNP	55134	564916	AGTGGGAGCATGTGGGAGACTCTTCTG>AGGTCCTGCTGACTGTTCTCT	G	C	C	C	T	G	G		
SNP	56683	568465	AATAAGGCCCAAGAAATAC>TTGAGACTGTCTTCCCGAGT	G	A	G	G	G	G	G		
SNP	56856	568638	CCAGGTGTGGTGGTGCCAGCGC>TTGTGATGCCAGCTCTCAGGAGCTGAG	C	T	C	C	C	C	C		
SNP	57790	567572	GGTTGGAGAAAGGGTGCTCTTCTTAT>CTATTTATTTTATTTTATTTATG	A	C	T	C	C	C	C		
SNP	60559	570341	TATTTGATTTGACTTAC>GATGAGAAATATCTAATATATTA	C	C	C	A	A	C	G		
SNP	60604	570386	CTATTTAATACAAATTTGCA>TTCAAGCTCCATAAACAATCC	A	C	C	G	G	A	T		
SNP	61165	570947	AGCTTTCTTGCCGCTGCG>GATGCTGCTGCTAGTCTGCTC	G	A	A	A	A	A	A		
SNP	64559	574341	CTGCATTTTGGGAAGCGGATG>CTTGAACACAGAGTCAGTATACAAAAT	C	C	C	C	C	T	C		
SNP	65171	574953	CATCCTTTAAAGAGAA>GACAGGTGTCTATATGTC	A	G	A	G	G	T	T		
SNP	65857	575939	AATTTAATTTGGCTAAGTACT>GTGACATCAGTTTGGGCCATTCAGTTTAT	T	G	T	T	T	T	T		
SNP	66184	575946	TCCCTTCAGGAAGCCAGAGGATCTCTG>CTTGAACACCAAA	C	C	C	C	T	T	T		
SNP	66190	575946	ACCACAAATTAAGGAATAAAGAGC>TCTTATATATCAGAAATTTCAATGTCT	C	C	C	T	T	T	T		
SNP	66526	576308	ACTGAGTATCTTGGCAGGCA>CTTATGTAATAAAGTATTAACATGA	C	T	C	C	C	C	C		
SNP	66902	576684	AAAATAAATATCTGTCTCCCATTA>GTGTATAGTGAATAGGAAGAATAA	A	A	A	A	A	A	A		
SNP	67857	577639	TAAATAAGTTTATTTGGAT>ATACAGCCAGGCCCATCC	T	T	T	T	T	A	A		
SNP	67919	577701	TTCCACTCTACATGGCAGAT>CTTTGAGTAGCTGTTA	T	T	T	T	T	C	T		
SNP	72270	582052	AATAGTAAATTAAGGAGGATGG>GTGAGTGAATTTGAGATTTCTGTA	G	A	G	G	G	G	G		
SNP	73115	584897	CAGTTGTTAAGGCGAGCTCGA>ATCACTGACTCTCTATGATGG	G	A	G	G	G	A	A		
SNP	76080	585962	GAGCCCATTCACCCCTTCCACAAC>TTACGAGGGCGAGGCA	C	C	C	C	C	C	C		
SNP	76101	585983	ACAACAGCGGGCGAGGCAACCCCTA>GTGAGCAATAAATAAT	G	A	G	G	G	C	C		
SNP	81912	591894	ACTGGTCAACATCTTGACACATATTA>TTTGGCGATTTCACTGG	A	G	T	T	T	T	T		
SNP	82203	591985	GAGCCACAATAGCAATGCTTAAGGAGC>GTAAAGCATCTCATTTTCAATG	A	G	A	A	G	A	G		
SNP	82332	592114	CTTTCACCTCTCAAAATGAATCAACG>CTTGGCTACTGCTCTTTTGTAAA	C	T	C	C	C	C	C		
SNP	82922	592704	TTGTATATATCTAATAGTAAATTT>ATTAATTTTTTCAATGGTTTTAT	T	T	T	T	T	T	A		
											AAA1 exon 4	no change (5'UTR)
											GPRA exon 3	Asp116

Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Polymorphism and flanking sequence	Alleles detected in haplotypes H1-H7							Location in GPRA and AAA1 exons	Amino acid change	
				H1	H2	H3	H4	H5	H6	H7			
SNP	83552	593334	AAGTGGCAACATATTATGAACATAGT(C>T)TCAAAGGTACACAGCATCAGTT	C	C	C	C	C	C	T	T	GPRA exon 4	no change
SNP	85227	595009	GCTCACCCATGTTTTTAACAT>CJACAGCCAGCACCATAGTAA	T	C	T	G	C	C/T	C	C		
SNP	85271	595053	AAGTAGTTAACACTCCAGCCCA(G>A)TGTTAAACTCTGCCTGCATGTGTC	G	G	G	G	G	G	A	A		
SNP	107610	617392	CCTAATCTTAAACCAATGAT>CJATTAAGAAAGAAAGAACTA	T	C	C	C	C	C/T	C	C		
SNP	110989	620771	AATGCCCTCTTAATTTTACTTT>CJCTTGATATACCTTTGGCCATCCCTCT	T	T	T	T	T	C	C	C		
SNP	111012	620794	CTTTCTTGATATACCTTTGGCCATCC(C>T)TCTATTCCACACTTAGATATCT	C	T	C	C	C	C	C	C		
SNP	112030	621812	TATCTATGCTCTCCCTCTCTT(G>C)JAATCTTAATTTTCAATCTC	G	G	G	G	G	C	C	C		
SNP	112037	621819	TCTCCCTCTCTGAAATCTT>GJAATTTAATTCAAATCTC	T	G	T	T	T	T	T	T		
SNP	112283	622065	TTCTTTAGGTAACTTTTAA(A>T)TTTTTACCCCTGCCCTTAA	A	T	A	T	A	T	T	T		
SNP	112726	622508	CCAAGAGATAATTTTCTTTT>ATAGTTTTCAGAGGAGCCTTC	T	T	T	A	A	A	T	T		
SNP	112859	622641	GTTTGTTGACAGATTATTTA>CJATCACCCAGGTATTTTTT	A	C	A	C	C	C	C	C		
SNP	113428	623210	TAATCATCTCTCAAGACCTCTC>GJCTATGTTGGCTTTCACACATTAGTG	C	C	C	A	C	C	G	G		
SNP	113845	623427	TCTACTCCCTCTGGAATTTGA(A>T)CCTAATGTTACAAAGTGA	T	T	A	T	A/T	T	T	T		
SNP	113944	623728	AGTGGATGAAAAGGGAGGCT>GJCTCTGATGGTGTGAGTGCCTG	T	T	T	T	G	T	T	T		
SNP	114945	624727	GGAATACCTGTTCTCATTTA>GJGACCACTGGTTTCCAGGCTT	A	A	A	C	A/G	A	C	A		
SNP	115192	624974	CATTTCTGCTTTAGGTTGTGCTG(C>G)TACGCCCTACCTACGTCCTGGT	C	G	C	C	C	A	C	C		
SNP	115628	625410	CATGCATACCTGTCCAAAGAAIT>CJGAACCTCCACACTAACTTCTCATTT	T	T	T	T	T	T	C	C		
SNP	116032	625814	GGGCCCTGGGAGCAGTAAT(G>A)CCCTAGAGGAGGCTGAGTA	G	G	G	A/G	G	G	G	G		
SNP	116464	626246	CTTATTGAATATAAATGT(G>A)CAGCCCTTTGATATGACTTC	T	T	T	T	A/G	A	G	G		
SNP	116515	626297	TCTGGAGAAACACCCACAC(G>A)TGTGCATGTATACATATGTG	T	G	G	G	G	A	A	A		
SNP	116926	626708	TAGAGCCATGGCAGTAAT>CJGTCATCAACATGAACATATC	T	T	T	T	T	T	C	C		
SNP	117276	627058	ATTATTGATCATGCTTGT>ATATGCTACTGTACAGCATAAGACATATG	T	T	T	T	T	T	A	A		
SNP	123667	633449	ACATTAAAGTGGTATGCAIT(G>A)ATATGATATAGCTTTTTTAT	G	A	G	A	A	A	G	A		
SNP	123770	633552	TGTTCTCGGGTTTCTCTCGGTC(A>G)TTTGGTGTCTGTTTGTGAAGGG	A	G	A	A	A	A	A	A		
SNP	123788	633570	GGTCATTGGTGTCTGTTT(G>C)TGAAGGGTTGAGACCTGTACCAAGTTT	G	C	G	G	G	G	G	G		
SNP	129017	638799	CAGAGAAAGAGGCCCTCCCTG(A>G)AGGTCACAGGGAGGGTGGAGGG	A	A	A	G	nd	G	G	nd		

no sequencing/genotyping result is available for the site

nd = no sequencing/genotyping result is available for the site

del=deletion

ins=insertion

Table 4. Primers used in the cloning of *GPRA*.

Primer	Sequence	Exon	Splice variant
JEGE1F1	tcgtgacctccgttcagcag	5' UTR	-
JEGE1F2	aagctggactccctcactcagc	5' UTR	-
JEGE1F3	agcaaggacagtgaggctcaacc	5' UTR	-
JEGE9aR1	ctggcatgaataaactggggaggtc	3' UTR	A
JEGE9aR2	tttgtcttgatcctccaggga	3' UTR	A
JEGE9bR1	tatagccctccctggatgaatctga	3' UTR	B
JEGE9bR2	gtgcccctggtaagcagtgagaagt	3' UTR	B
JEGE5F1	tgagcaattgataactctgtggctcctc	E2a	-
JEGExR2	AAATAAGCTGTGGCATCCTCATCCAG	3' UTR	C
JEGExR1	TGGCATCCTCATCCAGGATATTTGC	3' UTR	C
VAUE3R1	GGAAGGCCACGATGGTCATGTATGG	E5	-
Vau8R1	ATGACGAGGGTGGGGTGAAGTTGG	3' UTR	A
Vau8R2	GAATGGTGGGGAAGGAGCGGTTT	3' UTR	A
Vau1000R1	ggccatctctgtgacccatftt	3' UTR	B
AS8FX1	ATGAGATGCAGATTCTGTCCAAG	E9a	A
AS8FX2	TGCACAAGACAAAATGTTCTAATGA	3' UTR	A
AS8F5.1	CAGCTATAACCGAGGACTCATCTC	E7	-
AS8F7.1	TGTTGGAGTCCATCTCTCTGTT	E8	-
VAUE3R2	AAAAGACAGGCTCCAGGCGATCACA	E5	-
A2E2F	gattctcccagtggttgactgaa	E1	-
A3E3R	TGATGGCCAGCTGAGTCACAAAGAAGG	E2a	-

Table5. Primer pairs used in re-sequencing of the exons and exon/intron boundaries of *GPRA*

Primer	Forward primer	Reverse primer
<i>GPRA</i> ex1	actcagctgcaggagcaag	tgacactottaagtccagcagtc
<i>GPRA</i> ex2a	aggaggaagaaatccagcct	tgaccgatgcgttacatttt
<i>GPRA</i> ex2b.1	gccatacattgtagtaacctgaaa	cctcatccagggatatttgc
<i>GPRA</i> ex2b.2	ttcctaccaacaagaactccaa	cgatgatgaattagaacatacaacttt
<i>GPRA</i> ex3	taagtcaaagaactcctaccttgc	agcaaaggaaatacattaaaaatcaaa
<i>GPRA</i> ex4	ctgccctctttcaccagta	agccacccaccttccttagt
<i>GPRA</i> ex5	gcttctgtcaagctcccttt	gtgtggtcctgtcctgacg
<i>GPRA</i> ex6	tggatcctcatggtcactttc	tctctgctggcatagcttga
<i>GPRA</i> ex7	ttgagagagtctgagcattcca	ccaaattattcaaccatagcc
<i>GPRA</i> ex8	tgacatcaatgctccaacaa	tgcatgattaaggcggtttc
<i>GPRA</i> ex9a	ttaacatgtctacttgccitttca	tctgcaaaccgaggctatct
<i>GPRA</i> ex9b	gcagagctgtcacccaaaat	agcctgggcaacaagagtaa

Table 6A. Primers used in *GPRA* SNP genotyping.

SNP ID	Nucleotide	Type	Fw primer	Rev primer	Pe primer
NT 000380.3					
dbSNP: 324981	591694	A/T	GGCCATCTGATAAAGCAGGA	TGATGCATAGGAATGCAAGG	AGTCTCCAGTGAATCGCCAA

Table 6B. Primer pairs, restriction enzymes, and corresponding allele sizes used in genotyping of GPRA exonic SNPs. Six digit number in the name of the SNP shows its position in the genomic contig NT_000380.

Table 6B. Primer pairs, restriction enzymes, and corresponding allele sizes used in genotyping of GPRA exonic SNPs.							
Six digit number in the name of the SNP shows its position in the genomic contig NT_000380.							
Marker	Forward primer	Reverse primer	Enzyme	Product size*	Allele 1	Product size*	Allele 2
SNP_640764	aactgactaactaggtggccacgtgctccatcatcagaccatgctctt	tagcacaaatgctgccctat	SapI	340	t	290+50	c
SNP_662764	tgtctacctgttggcctgtg	gtctgtgcatctcccaggt	MboII	455	a	240+215	g
SNP_662803	catggagaaggaaaggtcagg	ctagcacitggcacitgcccta	DdeI	270	c	220+50	t
SNP_663133	atgactgcacgtcacttctta	tctgcaaacgcaggctatct	SfiNI	400	t	140+260	c
* product sizes are approximations							

* product sizes are approximations

Table 7. SNPs found in the exons of six splice variants of GPRA.

Exon	Position of the SNPs						Sequence around the polymorphism	Amino acid change in different variants
	seq1	seq2, A variant	seq4, B long variant	seq6, B short variant	seq8, C variant	seq10, D variant		
ex2b	-	-	-	-	585 -	seq14, F variant	TTTTCACTCCTATAA[C>T]CGTAGAAGTAGAG	no change (3'UTR of the C variant)
ex2b	-	-	-	-	655 -	seq12, E variant	CCTGGATGAGGATGCC[A>G]CAGCTTATTTCA	no change (3'UTR of the C variant)
ex2b	-	-	-	-	681 -	-	TTTTCATTATTTCTTC[G>A]ATTACAGTGTGGTAATG	no change (3'UTR of the C variant)
ex3	81912	448	448	415	-	448 -	TTGACAGATATT[A>T]TTGGCGATTCACT	Asn>Ile (A, B and E variants)
ex4	115192	524	524	491	-	420 -	CTGCTTTAGGTTGTGCTGCT[C>G]TACGCCCTCTACCTACGTTCC	no change (coding region of the A, B and D variants)
ex5	-	776	776	743	-	672 682	578 TACATGACCATCGTGGCCCTT [C>T]CTGGTGACTTTCATCCCTC *	no change (coding region of the A, B and F variants or in the 3'UTR of the D and E variants)
ex6	-	851	851	818	-	653	TATTTGGATTAAAG[C>G]AAACCTACGAAACAGT*	Ser>Arg (A, B and F variants)
ex8a	-	1159	-	-	-	861	CAGGGAGC[A>G]AAGATCACAGGATTCAGAAATG	Gln>Arg (A and F variants)
ex8a	-	1199	-	-	-	1001	GACGTTCCGGGAGAGAACT[T>C]GAGAGCATGAGATGCAGATTC*	no change (coding region of the A and F variants)
ex8a	-	1529	-	-	-	1331	CCAGTGAACACAGGCAAT[T>C]AGTGGTCCAGGTCCTGGCTT	no change (3'UTR of the A and F variants)
ex9b	-	-	1206	1173	-	-	CTAATGCTCTGCCCTCA[C>A]G AGAGAACTGG AAGGTA	no change (coding region of the B variants)
ex9b	-	-	1225	1192	-	-	AGAGAACTGGAGGGTA[C>T]TTGGCCAGGTGTACCTTCCTGG	Thr>Ile (B variants)
ex9b	-	-	1330	1297	-	-	TCTCACTGCTTACCAGGGCAC[A>T]GGACACC	no change (3'UTR of the B variants)
ex9b	-	-	1338	1305	-	-	GGACACC[A>G]GTGGTTCCCAAAATGGGTC	no change (3'UTR of the B variants)

* at these sites config NT_00380 contains an allele that associates with high serum IgE level and /or asthma.

Table 8. Best haplotype associations of *GPR4* for high serum IgE level in the data set of 304 disease associated and 220 control chromosomes when one gap (marked with an asterisks) was allowed in the haplotype patterns.

High IgE Control Conf χ^2				<i>G P R A</i> <i>m a r k e r s</i>				
				SNP591694 Exon 3	SNP640763 Exon 5	SNP662763 Exon 9A	SNP662803 Exon 9A	SNP663133 Exon 9A
51	16	0.761	10.3	-	C	-	-	-
44	14	0.759	8.5	T	C	-	-	-
40	12	0.769	8.5	-	C	*	T	-
39	12	0.765	7.9	-	C	*	T	C
39	12	0.765	7.9	-	C	G	-	-
39	12	0.765	7.9	-	C	G	T	-
38	12	0.76	7.3	-	C	G	*	C
38	12	0.76	7.3	-	C	G	T	C
35	11	0.761	6.8	T	C	*	T	-
34	11	0.756	6.2	T	C	*	T	C
34	11	0.756	6.2	T	C	G	-	-
34	11	0.756	6.2	T	C	G	T	-
33	11	0.75	5.7	T	C	G	*	C
33	11	0.75	5.7	T	C	G	T	C

High IgE, number of disease-associated haplotypes with a specific pattern; Control, number of control haplotypes with a specific pattern; Confidence, percentage of haplotypes with the specific pattern that is associate with disease, chi-square value for disease association of the specific haplotype pattern.

Table 9. Primers used for cloning of full length cDNAs for AAA1.

Primer	Exon	Sequence
IF	1	5' AGAATGAGTCTCTGATGACTTT 3'
IIF	2	5' ACTTGCTGTTTCATAGAATTGCAA 3'
SF9	4	5' TGACTTCTCCCCAGATTTTGTAT 3'
exIF	5	5' ACACATACAAAGTGCCTACCACAT 3'
SR13	9	5' TTGAAACTGTATTTCCCATATTGC 3'
ASKAF/R	10b	5' AAATGCAATAAAAAATGCGGAACTA 3'
XR	11	5' GAGTCATTAGTCCAGAGAACAT 3'
XIR	12	5' CTGCTTGGAACAGTGTATATC 3'
XIIR	14	5' TGGTCTACGTAGAATTCAGAGTA 3'
XVR	16	5' CATGTGTTAATTGTGTCTTCACT 3'
SR4	19	5' GGGTGTCAATTACACGAACAATAA 3'
SF10	6	5' CAGTTCAGTCACTGCAATCTTCAT 3'

Table 10. Primers used in SNP genotyping with SBE method.

SNP	forward	reverse	extension
517278	TTATGAGCTAAAGTGCCAAATTAAA	AGGAAGGTGGCAGTGAACTC	CCCTCCTATACATTACCTGAA
549709	TCCCGTCTCCTCTAGTCTTC	TGAGATGCTGTCTCTAAAATAAATAGA	TTCTCTCTTTGTACCCACAC
570341	TGTTGGTTTCGATGAGAGCAT	AAGTGGGGAACAACAACTGG	GTACCTATATTTGTATTGCACCTTA

Table 11. Exon-intron structures and splice junction sites of AAA1. Gray area shows the exons and introns located in AST1.

Exon	Size (bp)	site in NT_00380	5' splice donor	Intron size (kb)	5' splice acceptor
1	84	684692 - 684776	TACTGGAAAGgtatgattgt	37.0	gattttacagGCTGATGGTG
2	189	647354 - 647543	CTAAGACCTGgtaaagttaa	66.0	ctactgtgagGACACAGAGA
3	99	581535 - 581634	AGGTCTGCTGgtaaggaaat	7.0	ctattttcagAGTTCCTCAC
4	80	574305 - 574385	GAAAGCAAGGAGgtaaggagg	2.8	gctttacacagGAGTGACTTC
5	199	571268 - 571467	CATAGACCTGgtgagtaact	29.3	ctctggggagGTCTAGGACT
6	80	541930 - 542010	AAGAAAGCAAgtaagttag	5.3	ccattttacagCTTATGTGAG
7	112	536477 - 536589	CTCTGGTATGtaaggaaat	2.5	ggtttgacagGAGCTCCAAT
8	144	533835 - 533979	AAATAGCTGCAgttaagtaca	0.8	caattgtacagTGAACAAATA
9	947	532055 - 533002	TCCTCCATGAAggtctggac	14.7	catcttctagAAGTGAGAAA
10a	120	517273 - 517393	CTGGGTTCAggttaagta		caaaatcoagGAAAACAAGG
10b	351	517042 - 517393	TATGTCATHE	60.0	taaaaattagGTTAGGAAA
11	108	456437 - 456545	AGCTGGGTTGgtaagactgg	26.4	ttttttcagGTGGATATAC
12	97	430100 - 430197	CCATATATGAggtgatcata	47.2	cctttgcagGACTGGCAAC
13	150	382905 - 383055	TCTGCCATGTgtaagtcttc	1.3	gattctgaagGCCATGTTTT
14	121	381445 - 381566	GGGACATTTTgttttccatt	1.1	tgaatttttagGACGCTGGGC
15	71	380274 - 380345	ATGACCTGAGgtgggtat	0.3	ctgtccacagCAATCCTTCT
16	90	379916 - 380006	CGTTGTTACTaaaaaagaaa	149.0	tictcaacagAGCAAACTGC
17	94	230772 - 230866	TGAGCCTCAGgtaagaaccc	66.7	tgttttttagGCTTTTTGAT
18	118	168107 - 168224	TATTTTAAAGgtacatgat	4.1	tcttccccagCTCTGTTTTC
19	426	163615 - 164041	TTTCTAATTAAAAAaaaa		

Table 12. Best haplotype associations of AAA1 for high serum IgE level in the data set of 304 disease associated and 220 control chromosomes when one gap (marked with an asterisks) was allowed in the haplotype patterns. Only the markers from the coding region or close to intron-exon boundaries were included.

Markers of the AST1 locus							
High IgE	Control	X ²	SNP ₋₅₁₇₂₇₈ (Δ) Exon 10a / 10b	SNP ₋₅₃₈₅₆₇ (G>A) Intron 6	SNP ₋₅₄₉₇₀₉ (T>C) Intron 5	SNP ₋₅₇₀₃₄₁ (Δ) Intron 5	SNP ₋₅₇₄₉₅₃ (A>G) Intron 3
59/187 (24%)	18/179 (10%)	13.548	T	A	-	-	-
49/190 (21%)	14/177 (7.9%)	12.5479	T	A	*	G	-
49/190 (21%)	14/177 (7.9%)	12.5479	T	A	*	G	G
117/130 (47%)	55/181 (30%)	12.5318	T	-	-	-	-
53/189 (22%)	16/177 (9.0%)	12.2929	T	A	C	*	G
53/189 (22%)	16/177 (9.0%)	12.2929	T	A	C	-	-
47/180 (21%)	14/174 (8.0%)	12.2377	T	*	C	G	-
47/180 (21%)	14/174 (8.0%)	12.2377	T	*	C	G	G
47/190 (20%)	14/177 (7.9%)	11.4627	T	A	C	G	G
47/190 (20%)	14/177 (7.9%)	11.4627	T	A	C	G	-
49/190 (21%)	15/177 (8.4%)	11.3004	-	A	*	G	G
103/136 (43%)	47/174 (27%)	11.2636	T	*	C	-	-
50/187 (21%)	16/177 (9.0%)	10.9929	-	A	*	G	-
47/189 (20%)	15/177 (8.4%)	10.3762	-	A	C	G	-
47/190 (20%)	15/177 (8.4%)	10.2633	-	A	C	G	-
60/184 (25%)	23/179 (13%)	9.0248	-	A	-	-	G
53/189 (22%)	19/177 (11%)	8.957	-	A	C	*	-

High IgE, number of disease-associated haplotypes with a specific pattern, Control, number of control haplotypes with a specific pattern, chi-square value for disease association of the specific haplotype pattern

Table 13. Haplotype specific AST1 polymorphisms for diagnostic testing. These markers can be used for the haplotype identification of an individual without phase information.

Haplotype	Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Allele present in the specific haplotype	Allele present in other haplotypes
H1	SNP	4631	514413	T	C
H1	SNP	51975	561757	G	C
H1	SNP	53922	563704	T	C
H2	SNP	5634	515416	A	G
H2	SNP	6739	516521	A	G
H2	SNP	7550	517332	A	G
H2	deletion	9199-9201	518981-83	deletion of TCT	no deletion
H2	SNP	16980	526762	C	T
H2	SNP	17147	526929	A	G
H2	SNP	17435	527217	G	A
H2	SNP	19272	529054	T	A
H2	SNP	19452	529234	A	G
H2	SNP	20309	530091	A	G
H2	SNP	20789	530571	T	G
H2	SNP	24869	534651	C	T
H2	SNP	32976	542758	C	T
H2	SNP	34716	544498	C	A
H2	deletion	38850-52	548632-34	deletion of CTC	no deletion
H2	SNP	50955	560737	C	G
H2	SNP	51476	561258	G	A
H2	SNP	52573	562355	A	G
H2	SNP	55134	564916	A	G
H2	SNP	56856	566638	T	C
H2	SNP	65857	575639	G	T
H2	SNP	66526	576308	T	C
H2	SNP	66902	576684	G	A
H2	SNP	72270	582052	A	G
H2	SNP	111012	620794	T	C
H2	SNP	112037	621819	G	T
H2	SNP	115192	624974	G	C
H2	SNP	123770	633552	G	A
H2	SNP	123788	633570	C	G
H3	SNP	22869	532651	C	T
H4	deletion	7334-35	517116-17	deletion of AT	no deletion
H4	SNP	16893	526675	C	T
H4	SNP	17209	526991	A	C
H4	SNP	26198	535980	C/T*	C
H4	SNP	32124	541906	A/C*	C
H4	SNP	36551	546333	A	G
H4	SNP	116032	625814	A/G*	G
H6	SNP	54199	563981	C	T
H6	SNP	57790	567572	C	A
H6	SNP	64559	574341	T	C
H7	SNP	1	509783	T	C
H7	SNP	93	509875	G	A
H7	SNP	918	510700	A	G
H7	SNP	983	510765	T	A
H7	SNP	987	510769	C	T
H7	SNP	1542	511324	T	C
H7	SNP	1710	511492	G	A
H7	SNP	1818	511600	T	C
H7	SNP	1927	511709	T	A
H7	SNP	2254	512036	C	T
H7	SNP	4689	514471	G	C
H7	SNP	5442	515224	G	C
H7	SNP	18383	528165	A	G
H7	SNP	18978	528760	G	A
H7	SNP	19671	529453	A	G
H7	SNP	28866	538648	C	G
H7	SNP	52776	562558	C	G

* A rare polymorphism found only in H4 haplotype

Table 14. AST1 polymorphisms specific for the different haplotype combinations for diagnostic testing. These markers can be used for the identification of certain haplotype combinations of an individual without phase information.

Haplotype combination	Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Allele present in the specific haplotype combination	Allele present in other haplotypes
H4+H5	SNP	7125	516907	A	C
H4+H5	SNP	18927	528709	T	G
H4+H5	SNP	19268	529050	C	G
H4+H5	SNP	19712	529494	A	C
H4+H5	deletion	22122-23	531904-5	deletion of TG	no deletion
H4+H5	SNP	24007	533789	C	T
H4+H5	SNP	27404	537188	T	C
H4+H5	SNP	28785	538567	A	G
H4+H5	SNP	33350	543132	A	C
H4+H5	SNP	33798	543580	A	G
H4+H5	SNP	34382	544144	C	G
H4+H5	insertion	52286-87	562068-69	insertion of CC	no insertion
H4+H5	SNP	53803	563585	T	C
H4+H5	SNP	55000	564782	A	G
H4+H5	SNP	56683	566465	T	C
H4+H5	SNP	67919	577701	C	T
H4+H5	SNP	76101	585883	C	G
H4+H5	SNP	82332	592114	T	C
H4+H5	SNP	112726	622508	A	T
H4+H5	SNP	113944	623728	G	T
H4+H5	SNP	114945	624727	G	A
H4+H5	SNP	116464	626246	A	G
H4+H5	SNP	123667	633449	A	G
H2+H4+H5	SNP	4012	513794	A	C
H2+H4+H5	SNP	4961	514743	G	A
H2+H4+H5	SNP	5850	515632	G	A
H2+H4+H5	SNP	6312	516094	T	C
H2+H4+H5	SNP	6392	516174	T	G
H2+H4+H5	SNP	6485	516267	A	G
H2+H4+H5	SNP	6522	516304	G	C
H2+H4+H5	SNP	6646	516428	G	A
H2+H4+H5	SNP	6760	516542	C	T
H2+H4+H5	deletion	6821	516603	deletion of T	no deletion
H2+H4+H5	deletion	7240-43	517022-25	deletion of ACTT	no deletion
H2+H4+H5	SNP	7277	517059	G	C
H2+H4+H5	SNP	7303	517085	T	G
H2+H4+H5	SNP	7305	517087	C	G
H2+H4+H5	deletion	7308-8	517088-90	deletion of TGT	no deletion
H2+H4+H5	SNP	7496	517278	T	C
H2+H4+H5	SNP	8490	518272	T	C
H2+H4+H5	SNP	9649	518431	G	T
H2+H4+H5	deletion	9782-9785	519564-67	deletion of GTCT	no deletion
H2+H4+H5	SNP	11858	521640	G	A
H2+H4+H5+H7	SNP	3877	513659	C	G
H2+H4+H5+H7	SNP	7229	517011	T	C
H2+H4+H5+H7	SNP	10816	520598	C	T
H2+H4+H5+H7	SNP	12581	522363	C	G
H2+H4+H5+H7	SNP	19360	529142	A	G
H2+H4+H5+H7	SNP	19774	529556	A	C
H2+H4+H5+H7	SNP	20038	529820	C	T
H2+H4+H5+H7	SNP	20395	530177	C	T
H6+H7	SNP	22475	532257	C	T
H6+H7	SNP	22493	532275	G	A
H6+H7	SNP	22715	532497	A	G
H6+H7	SNP	22934	532716	T	A
H6+H7	SNP	24264	534046	T	G
H6+H7	SNP	26356	536138	T	C
H6+H7	SNP	26675	536457	G	A
H6+H7	SNP	28197	537979	A	G
H6+H7	SNP	28858	538640	C	T
H6+H7	SNP	31224	541006	A	G
H6+H7	SNP	32185	541967	C	T
H6+H7	deletion	34909	544691	deletion of T	no deletion

Table 14. (Continued)

Haplotype combination	Type of polymorphism	Position in SEQ ID NO:1	Position in NT_000380	Allele present in the specific haplotype combination	Allele present in other haplotypes
H6+H7	SNP	37327	547109	T	G
H6+H7	SNP	37415	547197	A	G
H6+H7	SNP	37685	547467	G	A
H6+H7	SNP	37959	547741	T	C
H6+H7	SNP	39343	549125	T	G
H6+H7	SNP	50493	560275	A	G
H6+H7	SNP	50835	560617	C	A
H6+H7	CA-repeat	51022-49	560804-31	(CA)8	(CA)6TA(CA)7
H6+H7	SNP	51536	561318	T	C
H6+H7	SNP	51884	561666	T	G
H6+H7	SNP	60604	570386	T	A
H6+H7	SNP	61165	570947	A	G
H6+H7	SNP	75115	584897	A	G
H6+H7	SNP	82922	592704	A	T
H6+H7	SNP	83552	593334	T	C
H6+H7	SNP	85271	595053	A	G
H6+H7	SNP	110989	620771	C	T
H6+H7	SNP	112030	621812	C	G
H6+H7	SNP	113428	623210	G	C
H6+H7	SNP	115628	625410	C	T
H6+H7	SNP	116515	626297	A	G
H6+H7	SNP	116928	626708	C	T
H6+H7	SNP	117276	627058	A	T
H2+H6+H7	SNP	28770	538552	T	C
H2+H6+H7	SNP	31910	541692	A	G
H2+H6+H7	SNP	37931	547713	T	C
H2+H6+H7	SNP	39314	549096	A	G
H2+H6+H7	SNP	50197	559979	T	G
H2+H6+H7	SNP	50334	560116	A	G
H2+H6+H7	SNP	50632	560414	A	G
H2+H6+H7	SNP	51217	560999	A	G
H2+H6+H7	SNP	51861	561643	G	C
H4+H5+H6+H7	SNP	54148	563930	A	G
H4+H5+H6+H7	SNP	54641	564423	C	G
H4+H5+H6+H7	SNP	54751	564533	C	T
H4+H5+H6+H7	SNP	60559	570341	G	C
H4+H5+H6+H7	SNP	66164	575946	T	C
H4+H5+H6+H7	SNP	67857	577639	A	T
H2+H7	SNP	16845	526627	C	T
H2+H7	SNP	20089	529871	A	T
H2+H4+H5+H6+H7	SNP	21850	531632	T	C
H2+H4+H5+H6+H7	SNP	36909	546691	A	C
H2+H4+H5+H6+H7	SNP	39927	549709	C	T
H2+H4+H5+H6+H7	SNP	45826	555608	T	C
H2+H4+H5+H6+H7	SNP	65171	574953	G	A
H2+H4+H5+H6+H7	SNP	112283	622065	T	A
H2+H4+H5+H6+H7	SNP	112859	622641	C	A

Table 15. The cellular location of transiently expressed GPRA A, B, B-short, C, D, E, and F variants.
Results are means at least from two independent experiments.

	GPRA A	GPRA B	GPRA B-short	GPRA C	GPRA D	GPRA E	GPRA F
cytoplasm	29%	48%	100%	100%	100%	100%	100%
Plasma membrane	71%	52%	0%	0%	0%	0%	0%

WHAT IS CLAIMED IS:

1. An isolated GPRA polypeptide comprising an amino acid sequence that has at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over the entire length of the selected SEQ ID NO. when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix.

2. The isolated GPRA polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of B-Long (SEQ ID NO: 5), B-Short (SEQ ID NO:7), C (SEQ ID NO:9), D (SEQ. ID NO:11), E (SEQ ID NO:13) and F (SEQ ID NO:15) provided that up to 34 amino acids can be substituted, deleted or inserted relative to the selected SEQ ID NO.

3. The isolated GPRA polypeptide of claim 1 comprising the amino acid sequence selected from the group consisting of B-Long (SEQ ID NO: 5), B-Short (SEQ ID NO:7), C (SEQ ID NO:9), D (SEQ. ID NO:11), E (SEQ ID NO:13) and F (SEQ ID NO:15)

4. An isolated GPRA polypeptide comprising at least 10 contiguous amino acids from amino acids 343-377 of B-long (SEQ ID NO:5).

5. An isolated GPRA polypeptide comprising an amino acid sequence that has at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over a sequence comparison window of at least 40 amino acids when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix provided that the polypeptide includes a variant amino acid encoded by a variant form shown in Table 7.

6. An isolated GPRA polypeptide of claim 1 comprising the amino acid sequence of SEQ ID NO:3 provided that the sequence contains an amino acid substitution of Asn 107Ile, Arg241Ser, or Gln344Arg.

7. An isolated GPRA polypeptide of claim 1 comprising the amino acid sequence of SEQ ID NO:5 provided that the sequence contain as an amino acid substitution of Asn instead of Ile at codon position 107, Arg instead of Ser at codon position 241, and/or Thr instead of Ile at codon position 366.

8. An isolated nucleic acid encoding the GPRA polypeptide of any of the preceding claims.

9. The isolated nucleic acid of claim 8 that hybridizes under highly stringent conditions to any of SEQ ID NOS: 4, 6, 8, 10, 12, and 14 wherein the highly stringent conditions are $6 \times \text{NaCl/sodium citrate (SSC)}$ at about 45°C for a hybridization step, followed by a wash of $2 \times \text{SSC}$ at 50°C

10. An isolated nucleic acid of claim 8 that hybridizes under highly stringent conditions to any of SEQ ID NOS: 1, 4, 6, 8, 10, 12, and 14 without hybridizing under the same highly stringent conditions to SEQ ID NO:2, wherein the highly stringent conditions are $6 \times \text{NaCl/sodium citrate (SSC)}$ at about 45°C for a hybridization step, followed by a wash of $2 \times \text{SSC}$ at 50°C .

11. An isolated nucleic acid of claim 8 having a sequence that is at least 90 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 4, 6, 8, 10, 12, and 14 over the entire length of the selected SEQ ID NO when compared using the BLASTN algorithm with a wordlength (W) of 11, $M=5$, and $N=-4$.

12. An isolated nucleic acid of claim 8 having a sequence that is at least 80 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 1, 2, 4, 6, 8, 10, 12, and 14 over a sequence comparison window of at least 100 nucleotides when compared using the BLASTN algorithm with a wordlength (W) of 11, $M=5$, and $N=-4$ provided that the nucleic acid includes a polymorphic site occupied by a variant form as shown in Table 3 or Table 7.

13. The isolated nucleic acid of claim 12, wherein the variant form is at a polymorphic site not designated by a * in Table 7.

14. An isolated nucleic acid of claim 8 having a sequence that is at least 80 % identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 1, 2, 4, 6, 8, 10, 12, and 14 over a sequence comparison window of at least 100 nucleotides when compared using the BLASTN algorithm with a wordlength (W) of 11, $M=5$, and $N=-4$ provided that the nucleic acid includes a polymorphic site occupied by a reference form designated with a * in Table 7.

15. An isolated genomic DNA molecule or a minigene having at least one intronic sequence and encoding a GPRA polypeptide that has at least 80 % sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13 and 15 over a region at least 40 amino acids in length when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix .
16. The isolated genomic DNA molecule or minigene of claim 15, comprising at least exons III-IV of SEQ ID NO: 2.
17. The isolated genomic DNA molecule or minigene of claim 15, wherein the intronic sequence is from intron 2 or 3 of a GPRA gene.
18. The isolated genomic DNA molecule or minigene of claim 15, wherein the GPRA polypeptide includes amino acids 343-377 of B-Long (SEQ ID NO:5)
19. The isolated genomic DNA molecule or minigene of claim 15, wherein the polypeptide is SEQ. ID NOS: 3, 5, 7, 9, 11, 13 or 15.
20. The isolated nucleic acid, genomic DNA molecule or minigene of any of claims 8-19, linked to a second nucleic acid with which it is not naturally associated.
21. The isolated nucleic acid, genomic DNA molecule or minigene of claim 20, wherein the second nucleic acid includes a heterologous promoter operably linked to a gene within the isolated nucleic acid.
22. A vector comprising the isolated nucleic acid, genomic molecule or minigene nucleic acid of any of claims 8-21.
23. A host cell comprising the vector of claim 22.
24. An antibody that specifically binds to an epitope within amino acids 343-377 of B-long (SEQ ID NO:5) or amino acids 332-366 of B-short (SEQ ID NO:7).
25. A method of preventing or treating asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer, comprising
administering to a patient suffering from or at risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer an effective amount of a

modulator of a GPRA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13, and 15.

26. The method of claim 25, further comprising administering an effective amount of a modulator of a AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

27. The method of claim 25, wherein the modulator binds to the GPRA polypeptide.

28. The method of claim 25, wherein the modulator inhibits expression of the GPRA polypeptide.

29. The method of claim 25, wherein the modulator is a transcript of an AAA1 gene or an inhibitor of expression of an AAA1 gene.

30. The method of claim 26, wherein the modulator of the AAA1 polypeptide binds to the AAA1 polypeptide.

31. The method of claim 26, wherein the modulator of the AAA1 polypeptide inhibits expression of the AAA1 polypeptide.

32. A method of identifying a modulator of a GPRA polypeptide, comprising

contacting a cell expressing a GPRA polypeptide with an agent;
determining whether the agent modulates expression of the GPRA polypeptide and/or signal transduction through the GPRA polypeptide, wherein the GPRA polypeptide is defined by any of claims 1-7.

33. The method of claim 32, wherein the cell further expresses an AAA1 polypeptide.

34. A method of determining risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer, comprising;

determining whether the individual has a variant polymorphic form in a GPRA gene, wherein presence of the variant polymorphic form indicates risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

35. The method of claim 34, wherein the variant form occurs in a noncoding region of the GPRA gene

36. The method of claim 34, wherein the variant form occurs in a coding region of the GPRA gene

37. The method of claim 34, wherein the variant form occurs between introns 1 and 4 of the GPRA gene.

38. The method of claim 34, wherein the determining comprises determining whether the individual has a variant form relative to SEQ ID NO: 1 (AST-1 locus).

39. The method of claim 34, wherein the determining comprises determining whether the individual has a variant form relative to any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, and 15.

40. The method of claim 34, wherein the variant form is a variant form shown in Table 3.

41. The method of claim 40, wherein the variant form is at a polymorphic site shown in Table 3.

42. The method of claim 34, wherein the variant form is a variant form shown in Table 7.

43. The method of claim 34, wherein the variant form is a variant form at a polymorphic site not designated * in Table 7.

44. The method of claim 34, wherein the determining comprising determining whether the individual has variant polymorphic forms relative to SEQ ID NO:1 at each of a plurality of polymorphic sites within the AST-1 locus, the presence of variant polymorphic forms at two or more of the plurality of polymorphic sites indicating increased risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

45. The method of claim 34, further comprising determining whether the individual has a variant polymorphic form in an AAA1 gene, wherein presence of the variant

polymorphic form indicates risk of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer.

46. The method of claim 45, wherein the variant polymorphic form occurs in the coding region of the AAA1 gene.

47. The method of claim 34, further comprising amplifying at least part of SEQ ID NO:1 (AST-1) locus including the polymorphic site before the determining step.

48. The method of claim 34, wherein the determining is performed by allele specific amplification, allele specific hybridization, single strand conformation polymorphism (SSCP), oligonucleotide ligation assay, single-base extension assay, or restriction fragment length polymorphism (RFLP).

49. A method for identifying a polymorphic site correlated with a disease selected from the group consisting of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease and cancer or susceptibility thereto, comprising;

identifying a polymorphic site within a GPRA gene,

determining whether a variant polymorphic form occupying the site is associated with the disease or susceptibility thereto.

50. The method of claim 49, wherein the variant form occurs in a noncoding region of the GPRA gene

51. The method of claim 49, wherein the variant form occurs in a coding region of the GPRA gene

52. The method of claim 49, wherein the variant form occurs between introns 2 and 4 of the GPRA gene.

53. The method of claim 49, wherein the determining comprises determining whether the individual has a variant form relative to SEQ ID NO: 1 (AST-1 locus)

54. The method of claim 49, wherein the determining comprises determining whether the individual has a variant form relative to any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, and 15.

55. The method of claim 49, wherein the determining is performed by comparing the frequency of the variant polymorphic form in individuals with and without the disease.

56. A primer or probe nucleic acid that hybridizes under highly stringent conditions to a segment of SEQ ID NO:1, 2 or 4 or a variant form thereof differing from SEQ ID NO: 1, 2 or 4 at a position shown in Table 3 or Table 7, wherein the segment includes or is immediately adjacent to a polymorphic site shown in Table 3 or Table 7.

57. The primer or probe of claim 56, wherein the position is a position other than a position designated * in Table 7.

58. The primer or probe of claim 56 that is perfectly complementary to a segment of SEQ ID NO:1, 2 or 4.

59. The primer or probe of claim 56 that is perfectly complementary to a variant form of a segment of SEQ ID NO:1, 2 or 4 shown in Table 3 or Table 7.

60. The primer or probe of claim 56 that specifically hybridizes to the segment of SEQ. ID NO: 1, 2 or 4 without hybridizing to a corresponding segment of an allelic variant shown in Table 3 or Table 7.

61. A primer of claim 56 for conducting a single-base extension reaction, whereby the primer is perfectly complementary to a segment that is immediately adjacent to but does not include the polymorphic site.

62. A transgenic animal comprising a nucleic acid according to any one of claims 8-21.

63. The transgenic animal of claim 62, further comprising a nucleic acid encoding an AAA1 polypeptide.

64. A transgenic animal of claim 62 disposed to develop a characteristic of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer in

which an endogenous a GPRA gene encoding a cognate form of a GPRA polypeptide defined by any of SEQ ID NOS: 3, 5, 7, 9, 11, 13 and 15 is functionally disrupted to prevent expression of a gene product.

65. The transgenic animal of claim 64 in which an endogenous AAA1 gene encoding a cognate form of an AAA1 polypeptide defined by an of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 is functionally disrupted to prevent expression of a gene product of the AAA1 gene.

66. A kit for use in diagnosing or assessing predisposition to asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer, comprising;

a container; and in the container:

a compound, preferably labeled, capable of detecting a polymorphic form at a polymorphic site in a susceptibility locus for asthma as defined by SEQ ID NO:1, 2 or 4.

67. The kit according to claim 66, wherein the polymorphic site occurs at a position shown in Table 3, Table 7, Table 12, Table 13 or Table 14.

68. The kit according to claim 66, wherein the compound is a primer or probe.

69. The kit according to claim 66, wherein said primer is the primer of claim 56.

70. An isolated AAA1 polypeptide comprising an amino acid sequence that has at least 80% sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 over the entire length of the selected SEQ ID No: when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix.

71. The isolated AAA1 polypeptide of claim 70 comprising an amino acid sequence that has at least 90% sequence identity to the selected amino acid sequence.

72. The isolated AAA1 polypeptide of claim 70 comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

73. The isolated AAA1 polypeptide of claim 70 comprising at least 10 contiguous amino acids from an amino acid sequence selected from the group consisting of SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

74. An isolated nucleic acid encoding the AAA1 polypeptide of any of claims 70-73.

75. The isolated nucleic acid of claim 74 that hybridizes under highly stringent conditions to any of SEQ ID NOS: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40, wherein the highly stringent conditions are $6 \times$ NaCl/sodium citrate (SSC) at about 45 °C for a hybridization step, followed by a wash of $2 \times$ SSC at 50 °C

76. An isolated nucleic acid of claim 74 having a sequence that is at least 80% identical to a nucleic acid having a sequence selected from the group consisting of SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 over the entire length of the selected SEQ ID NO when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N= -4.

77. The isolated nucleic acid of claim 76, wherein at least one polymorphic site shown in Table 12 is occupied by a variant nucleotide.

78. An isolated nucleic acid having at least 20 contiguous nucleotides from a sequence selected from the group consisting of SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40.

79. An isolated genomic DNA molecule or a minigene having at least one intronic sequence and encoding an AAA1 polypeptide that has at least at least 80 % sequence identity to an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 over the entire length of the selected SEQ ID NO. when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix .

80. The isolated genomic DNA molecule or minigene of claim 79, wherein the polypeptide has a sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

81. The isolated nucleic acid, genomic DNA molecule or minigene of claim 79 or 80 linked to a second nucleic acid with which it is not naturally associated.

82. The isolated nucleic acid, genomic DNA molecule or minigene of claim 79, wherein the second nucleic acid includes a heterologous promoter operably linked to a gene within the isolated nucleic acid.

83. A vector comprising the isolated nucleic acid, genomic molecule or minigene nucleic acid of any of claims 74-82.

84. A host cell comprising the vector of claim 83.

85. An antibody that specifically binds to a polypeptide selected from the group consisting of SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

86. A method of preventing or treating asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer, comprising administering to a patient suffering from or at risk of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer an effective amount of a modulator of an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

87. The method of claim 86, wherein the modulator binds to the AAA1 polypeptide.

88. The method of claim 86, wherein the modulator inhibits or induces expression of the AAA1 polypeptide.

89. A method of identifying a modulator of an AAA1 polypeptide, comprising

contacting an AAA1 polypeptide with an agent;

determining whether the agent binds to the AA1 polypeptide, modulates expression of the AAA1 polypeptide or modulates activity of the AAA1 polypeptide, wherein the AAA1 polypeptide comprises an amino acid sequenced as defined by any of SEQ ID NOS:17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41.

90. The method of claim 89, wherein the AAA1 polypeptide is expressed from a cell.

91. A method of determining risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer, comprising;
determining whether the individual has a variant polymorphic form in an AAA1 gene, wherein presence of the variant polymorphic form indicates risk of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

92. The method of claim 91, wherein the variant form occurs in a noncoding region of the AAA1 gene

93. The method of claim 91, wherein the variant form occurs in a coding region of the AAA1 gene

94. The method of claim 91, wherein the determining comprises determining whether the individual has a variant form relative to SEQ ID NO: 1 (AST-1 locus)

95. The method of claim 91, wherein the determining comprises determining whether the individual has a variant form relative to any of SEQ ID NOS: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40.

96. The method of claim 91, wherein the variant form is a variant form shown in Table 12.

97. The method of claim 91, wherein the determining is performed by allele specific amplification, allele specific hybridization, single strand conformation polymorphism (SSCP), oligonucleotide ligation assay, single-base extension assay, or restriction fragment length polymorphism (RFLP).

98. A method for identifying a polymorphic site correlated with a disease selected from the group consisting of asthma, other IgE mediated disease, chronic obstructive pulmonary disease or cancer or susceptibility thereto, comprising;

identifying a polymorphic site within an AAA1 gene,
determining whether a variant polymorphic form occupying the site is associated with the disease or susceptibility thereto.

99. The method of claim 98, wherein the variant form occurs in a noncoding region of the AAA1 gene.
100. The method of claim 98, wherein the variant form occurs in a coding region of the AAA1 gene
101. The method of claim 98, wherein the determining comprises determining whether the individual has a variant form relative to SEQ ID NO: 1 (AST-1 locus)
102. The method of claim 98, wherein the determining comprises determining whether the individual has a variant form relative to any of SEQ ID NOS: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40.
103. The method of claim 98, wherein the determining is performed by comparing the frequency of the variant polymorphic form in individuals with and without the disease.
104. A primer or probe nucleic acid of nucleotides that hybridizes under highly stringent conditions to a segment of SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 or a variant form thereof differing from SEQ ID NOS.16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40 at a single polymorphic site.
105. The primer or probe of claim 104, wherein the polymorphic site is one shown in Table 3 or 12.
106. The primer or probe of claim 104 that is perfectly complementary to a segment of SEQ ID NOS:16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38 and 40.
107. A primer or probe of claim 104 for conducting a single-base extension reaction, whereby the primer is perfectly complementary to a segment that is immediately adjacent to but does not include the polymorphic site.
108. A transgenic animal comprising a nucleic acid according to any one of claims 74-82.
109. A transgenic animal of claim 108 disposed to develop a characteristic of asthma, other IgE-mediated disease, chronic obstructive pulmonary disease or cancer in

which an endogenous a AAA1 gene encoding a cognate form of an AAA1 polypeptide defined by any of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 and 41 is functionally disrupted to prevent expression of a gene product.

110. A pharmaceutical composition comprising an effective amount of a GPRA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13, and 15 or a modulator thereof for the prevention or treatment of asthma or other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

111. The pharmaceutical composition of claim 110, further comprising an effective amount of an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and 41 or a modulator thereof.

112. The pharmaceutical composition of claim 110, wherein the modulator binds to the GPRA polypeptide and/or activates/inhibits expression of the GPRA polypeptide.

113. The pharmaceutical composition of claim 111, wherein the modulator is a transcript of an AAA1 gene or an activator/inhibitor of expression of an AAA1 gene.

114. The pharmaceutical composition of claim 111, wherein the modulator binds to the AAA1 polypeptide and/or activates/inhibits expression of the AAA1 polypeptide.

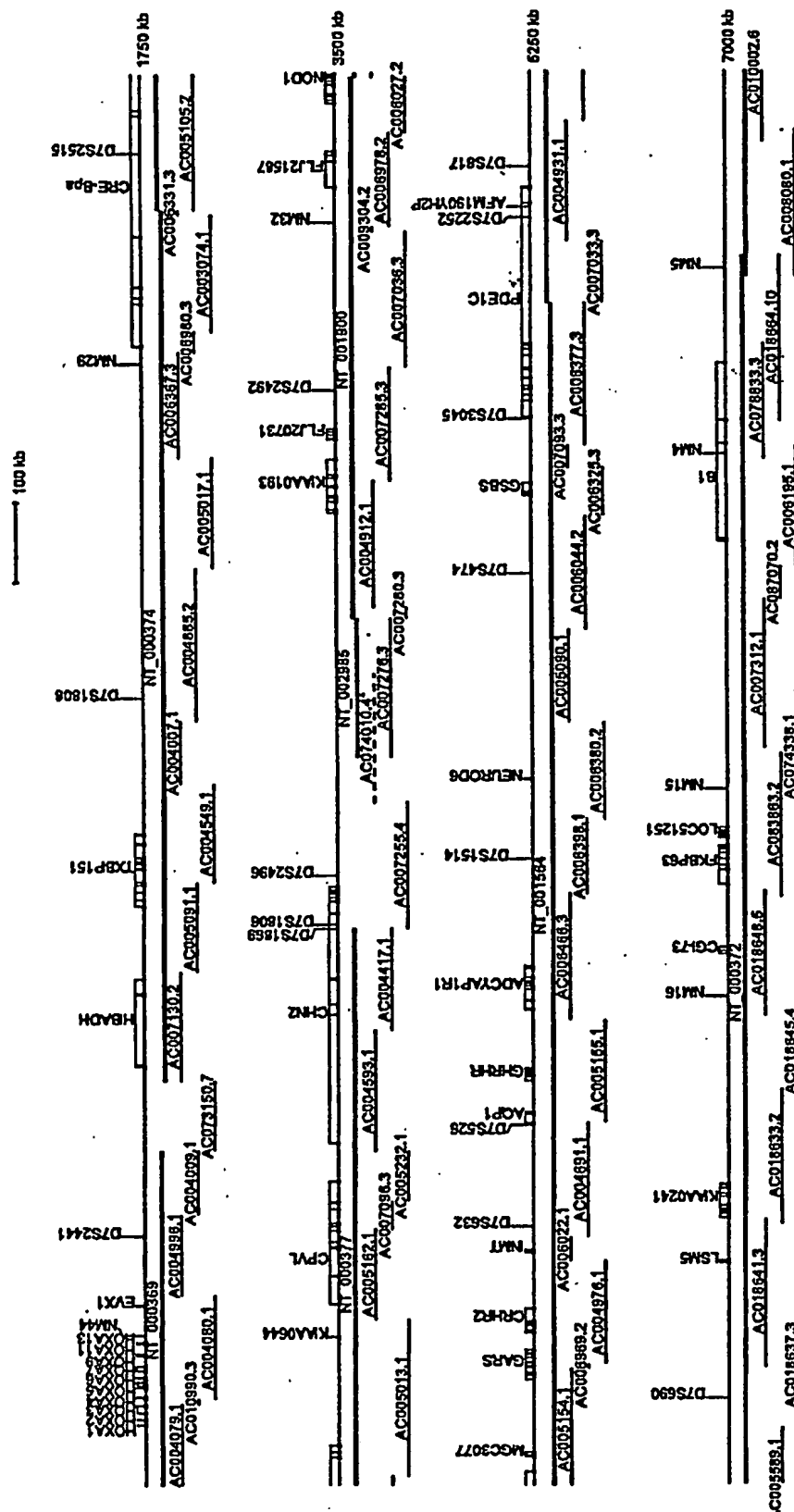
115. Use of a GPRA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, and 13 or a modulator thereof for the manufacture of a medicament for the prevention or treatment of asthma or other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

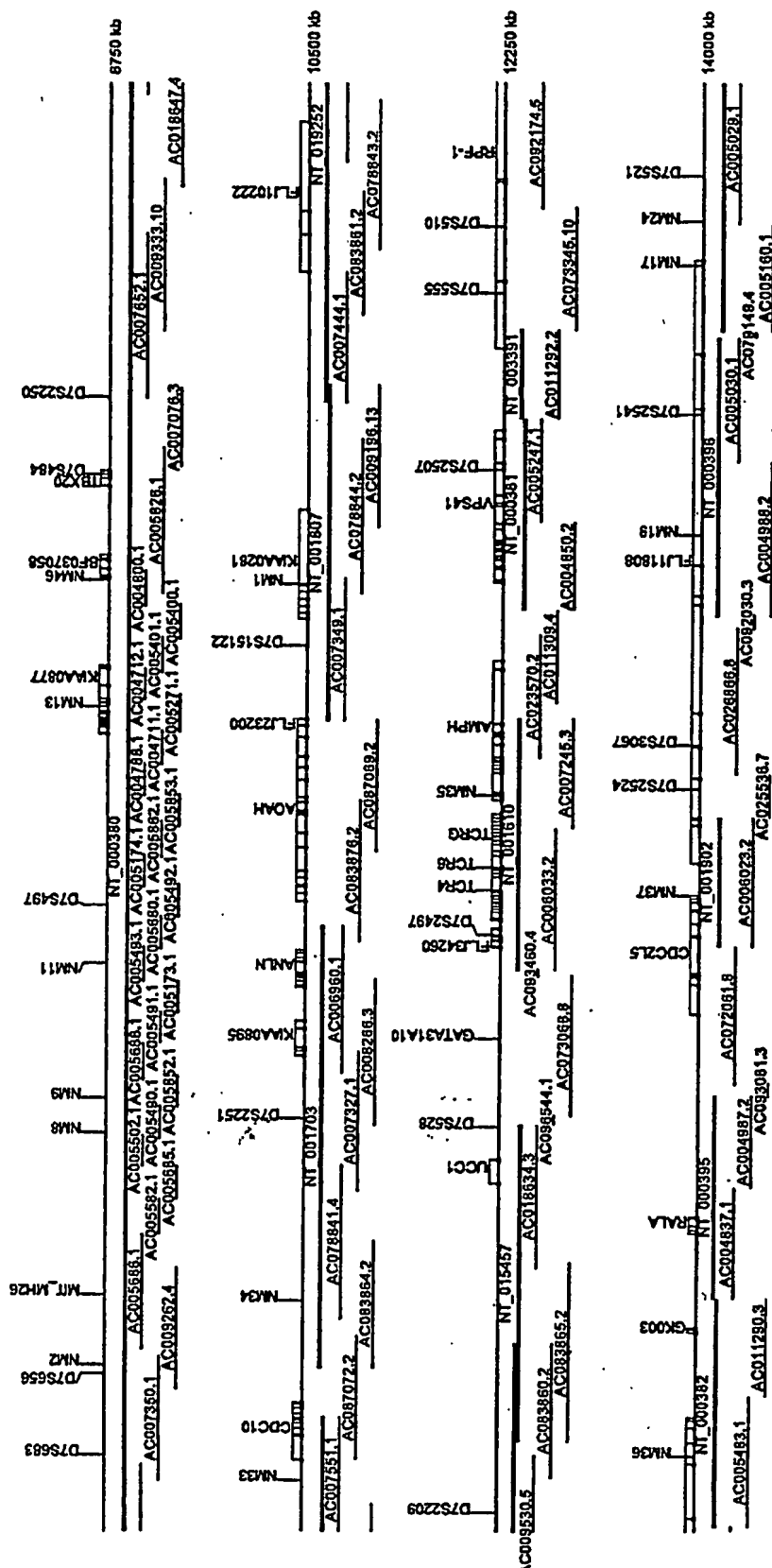
116. The use of claim 115, wherein a GPRA polypeptide comprising an amino acid sequence selected from the group consisting of SEQ. ID NOS: 3, 5, 7, 9, 11, 13, and 15 or a modulator thereof, and an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and 41 or a modulator thereof are used for the manufacture of said medicament.

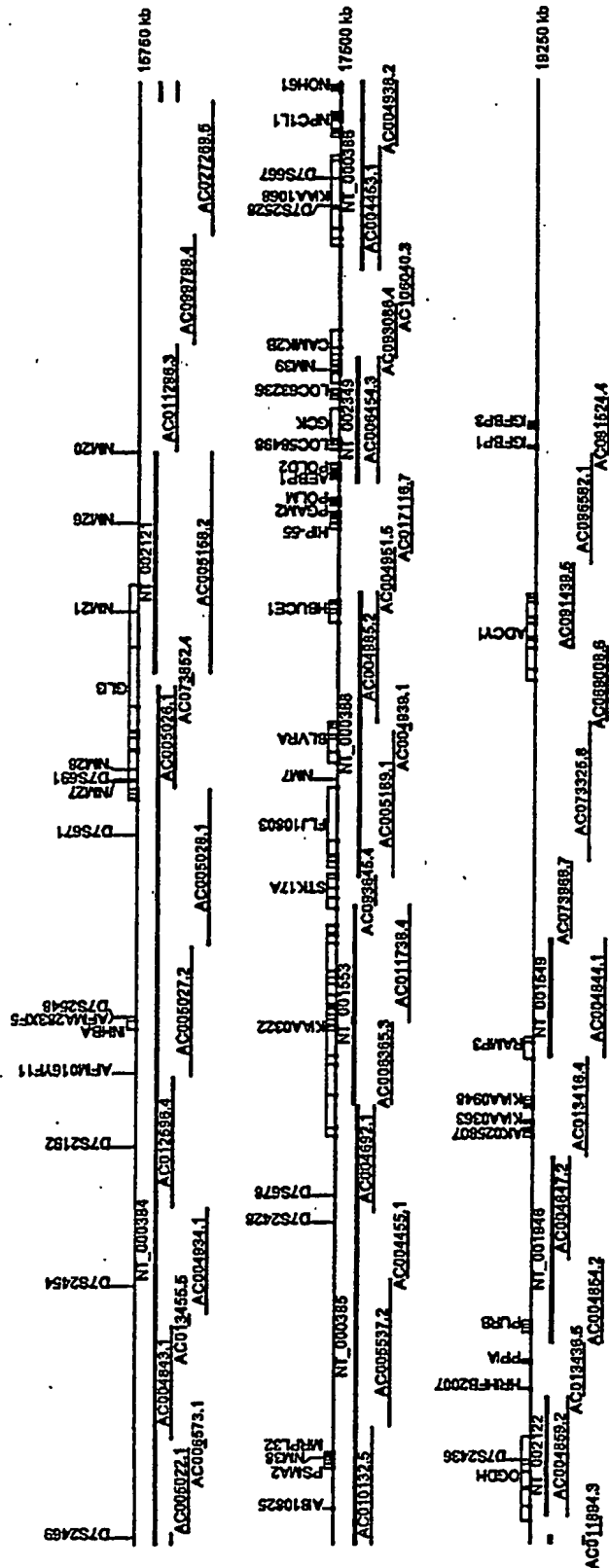
117. A pharmaceutical composition comprising an effective amount of an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and 41 or a modulator thereof for the prevention or treatment of asthma or other IgE mediated disease, chronic obstructive pulmonary disease or cancer.

118. The pharmaceutical composition of claim 117, wherein the modulator binds to the AAA1 polypeptide and/or activates/inhibits expression of the AAA1 polypeptide.

119. Use of an AAA1 polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, and 41 or a modulator thereof for the manufacture of a medicament for the prevention or treatment of asthma or other IgE mediated disease, chronic obstructive pulmonary disease or cancer.







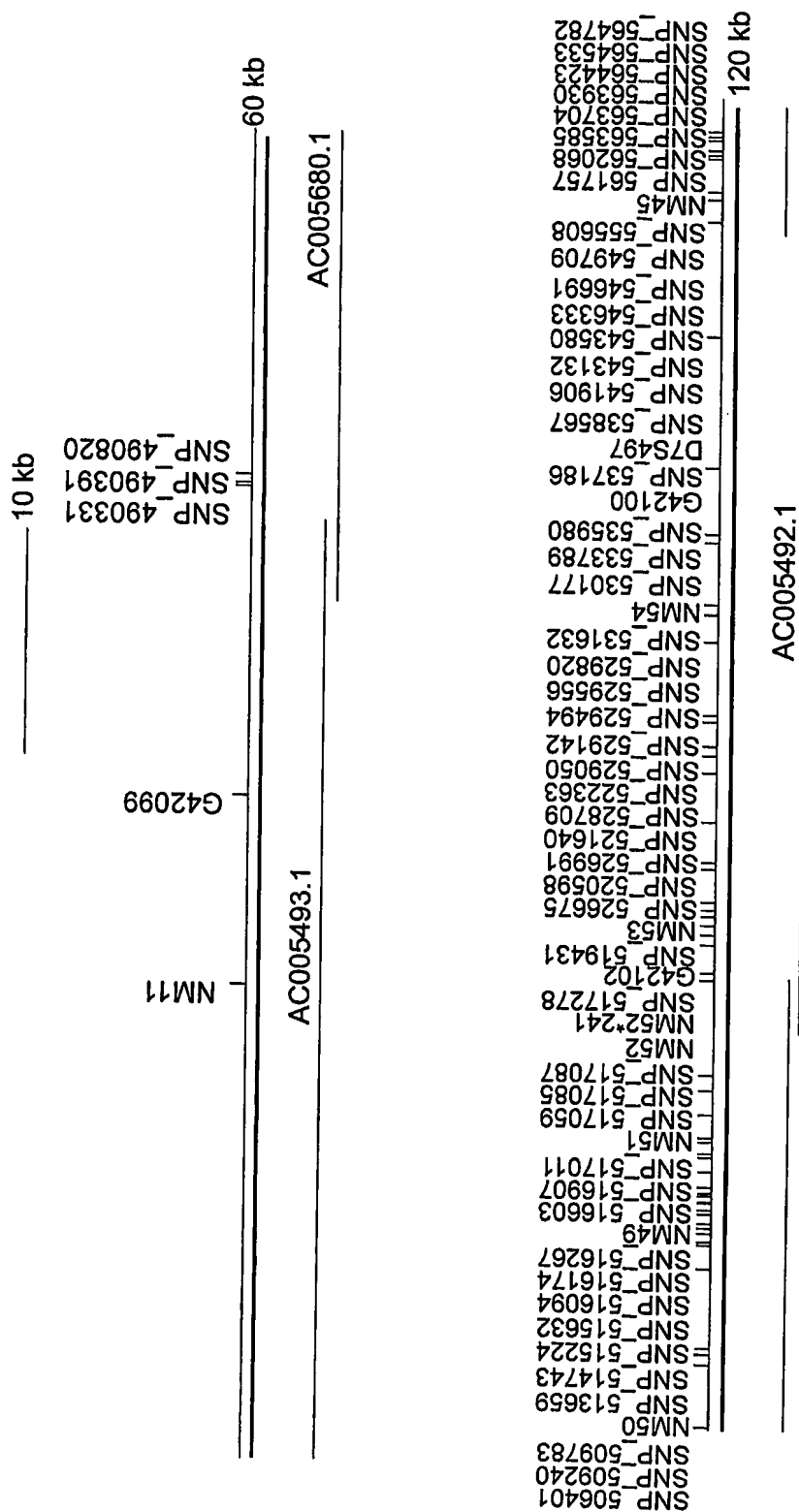


Fig. 2

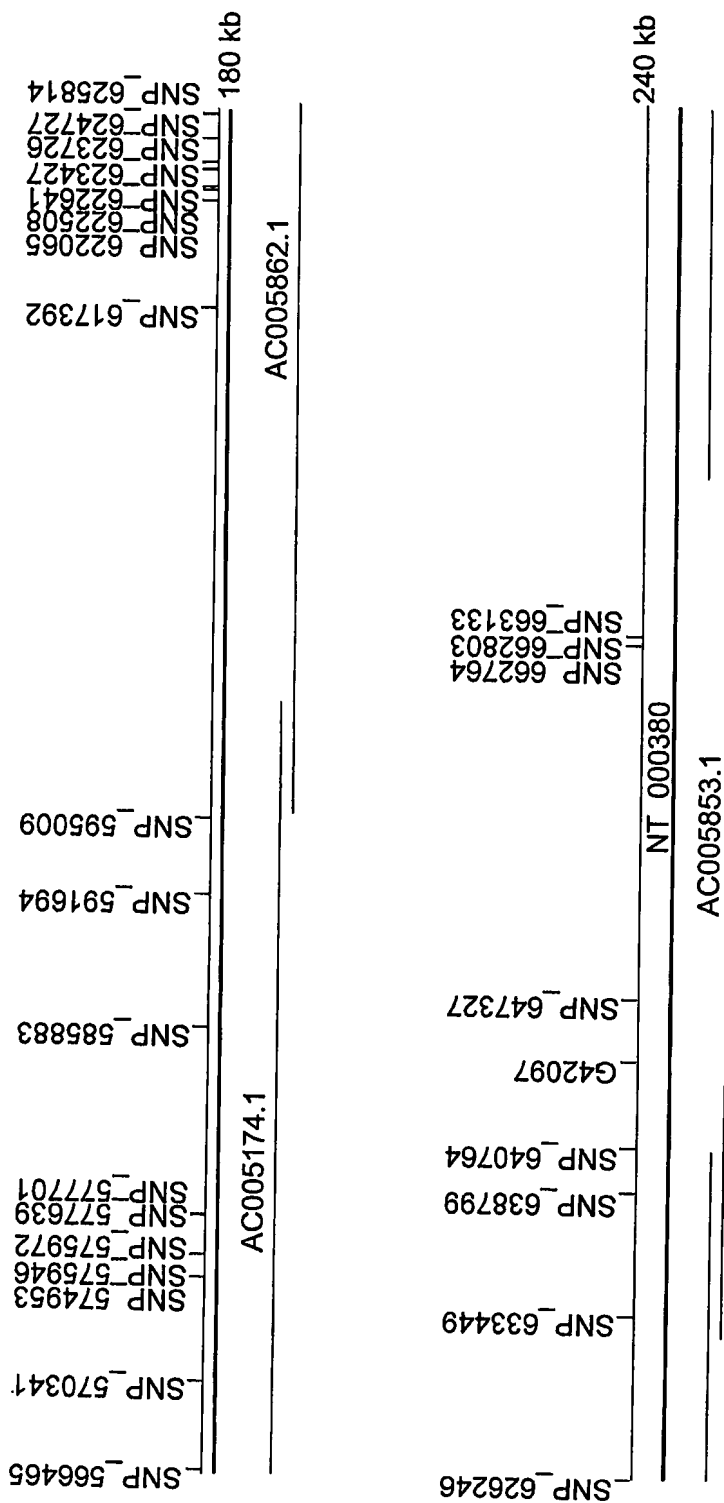


Fig. 2 (Continued)

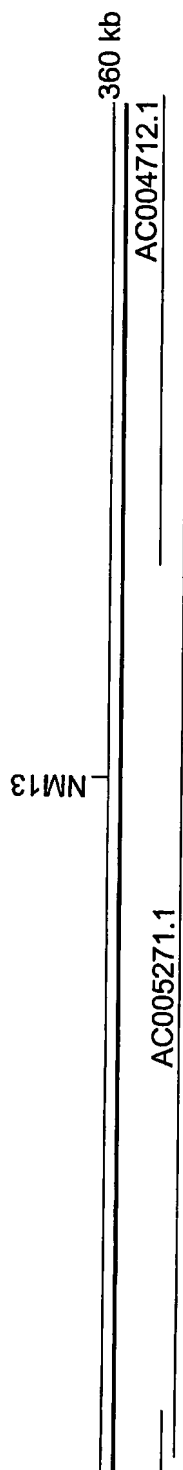


Fig. 2 (Continued)

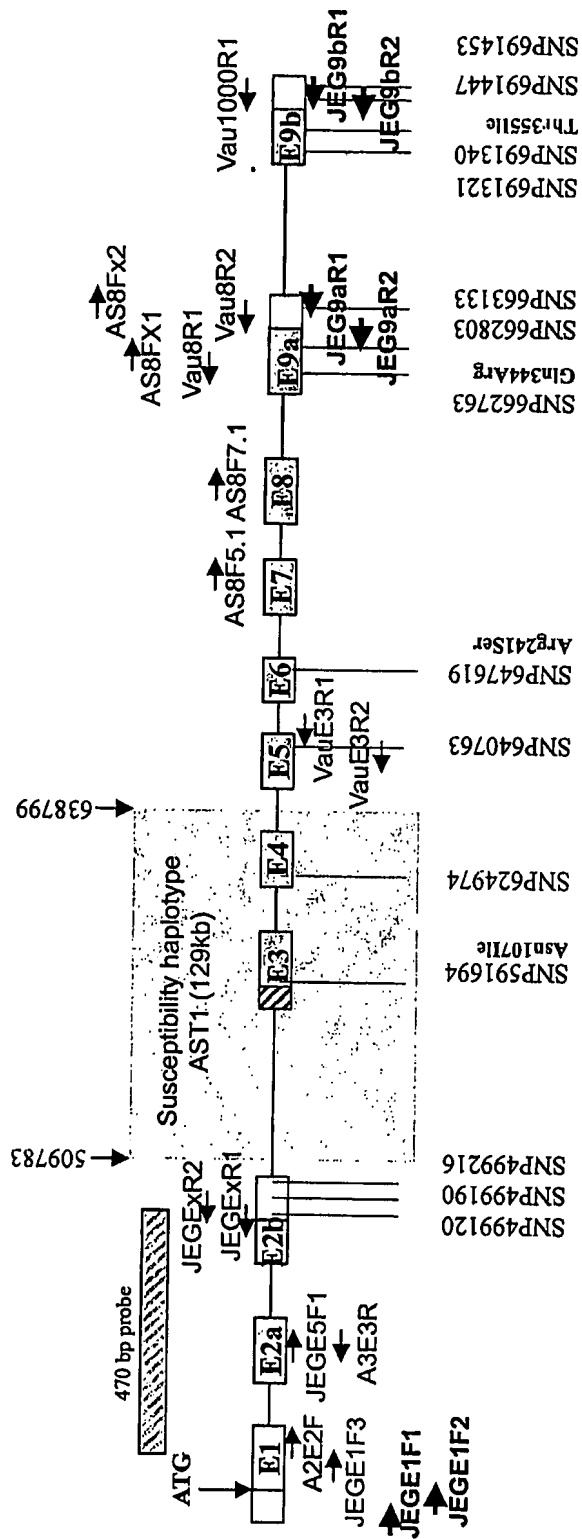


Fig. 3

Figure 4A**A SEQ ID NOs: 2 and 3 (GB AY310326)**

ggagggtcctgtgcctccgttcagcagagctgcagctgctgccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgagggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgaggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgactgaaacagtgacttttactgaagtggaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctctt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttggaactccgttggtgtttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
taagaatgaccttcttctgtgactcagctggccatcacag|attctttcacaggactggtc
S R M T F F V T Q L A I T D S F T G L V
aacatcttgacagatattaattggcgattcactggagacttcacggcacctgacctggtt
N I L T D I N W R F T G D F T A P D L V
tgccgagtggtccgctatttgag|gttggtgctgctctacgcctctacctaagtcctggtg
C R V V R Y L Q V V L L Y A S T Y V L V
tccctcagcatagacagataccatgccatcgtctaccccatgaagtcccttcaaggag|aa
S L S I D R Y H A I V Y P M K F L Q G E
aagcaagccagggtcctcattgtgatcgccctggagcctgtcttttctgttctccattccc
K Q A R V L I V I A W S L S F L F S I P
accctgatcatatttggaagaggacactgtccaacgggtgaagtgcagtgctgggacctg
T L I I F G K R T L S N G E V Q C W A L
tggcctgacgactcctactggacccatacatgaccatcgtggccttcctggtgtacttc
W P D D S Y W T P Y M T I V A F L V Y F
atccctctgacaatcatcag|catcatgtatggcattgtgatccgaactatttgattaaa
I P L T I I S I M Y G I V I R T I W I K
agcaaaacctacgaaacagtgatttccaactgctcag|atgggaaactgtgcagcagctat
S K T Y E T V I S N C S D G K L C S S Y
aacgaggactcatctcaaaggcaaaaatcaaggctatcaagtatagcatcatcatcatt
N R G L I S K A K I K A I K Y S I I I I
cttg|ccttcatctgctgttgaggatccatacttctgtttgacattttggacaatttcaac
L A F I C C W S P Y F L F D I L D N F N
ctccttccagacacccaggagcgtttctatgcctctgtgatcattcagaacctgccagca
L L P D T Q E R F Y A S V I I Q N L P A
ttgaatagtgcctcaacccctcatctactgtgtcttcagcagctccatctctttcccc
L N S A I N P L I Y C V F S S S I S F P
tgag|ggagcaaagatcacaggattccagaatgacgttccgggagagaactgagaggcat
C R E Q R S Q D S R M T F R E R T E R H
gagatgcagattctgtccaagccagaattcatctagaccctaggggcagtgccagtgtctag
E M Q I L S K P E F I *
gctgagcaccatcagctctcccaggtccttgacactgcttgggcacgtgcatggaaccc
gagccaacttcacccaccctcgtcattacctgggagatgcacaagacaaatgttcta
gactgcatgcaactgttaagtattggccaacacgaactccccagttattcatgccagcca
ggaaggaaacgccttcttccccaccattcccagccctccttccactggccagcacctg
aaccagtgaaacaggcatttagtgggtccagggctcctgggttgagccagtgagtagac

9/42

Figure 4B1**B long SEQ ID NOS: 4 and 5 (GB AY310327)**

ggggtcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgagggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgaggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggtgcactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctcttt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttgtgctttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaaatgacacttcttctgtgactcagctggccatcacag|attctttcacaggactggtc
S R M T F F V T Q L A I T D S F T G L V
aacatcttgacagatattaattggcgattcactggagacttcacggcacctgacctggtt
N I L T D I N W R F T G D F T A P D L V
tgccgagtggtccgctatattgag|gttggtgctgctctacgcctctacctacgtcctgggtg
C R V V R Y L Q V V L L Y A S T Y V L V
tccctcagcatagacagataccatgccatcgtctaccccatgaagttccttcaaggag|aa
S L S I D R Y H A I V Y P M K F L Q G E
aagcaagccagggtcctcattgtgatcgccctggagcctgtcttttctgttctcattccc
K Q A R V L I V I A W S L S F L F S I P
accctgatcatatttggaagaggacactgtccaacgggtgaagtgcagtgctgggccctg
T L I I F G K R T L S N G E V Q C W A L
tggcctgacgactcctactggaccccatatgaccatcgtggccttctggtgtacttc
W P D D S Y W T P Y M T I V A F L V Y F
atccctctgacaatcatcag|catcatgtatggcattgtgatccgaactatttgattaaa
I P L T I I S I M Y G I V I R T I W I K
agcaaacctacgaaacagtgatttccaactgctcag|atgggaaactgtgcagcagctat
S K T Y E T V I S N C S D G K L C S S Y
aaccgaggactcatctcaaaggcaaaaatcaaggctatcaagtatagcatcatcatcatt
N R G L I S K A K I K A I K Y S I I I I
cttg|ccttcatctgctgttgaggatccatacttccctgtttgacattttggacaatttcaac
L A F I C C W S P Y F L F D I L D N F N
ctccttccagacacccaggagcgtttctatgcctctgtgatcattcagaacctgccagca
L L P D T Q E R F Y A S V I I Q N L P A
ttgaatagtgccatcaacccccctcatctactgtgtcttcagcagctccatctcttcccc
L N S A I N P L I Y C V F S S S I S F P
tgcag|ggatcatccgtctccgtcagctccaggaggctgcgctaagtctctgccctcaacga
C R V I R L R Q L Q E A A L M L C P Q R
gagaactggaagggtacttggccagggtgtaccttccctgggtcttccaagggtgacagctc
E N W K G T W P G V P S W A L P R *
tcacctgtgctgcaggtggccctgtgcctgggtgccacttctcactgcttaccagggcac
aaggacaccagtggttcccaaatgggtcacagcaggatggcctgcatcagattcaccag
ggagggctataagaaggcagac

10/42

Figure 4B2**B short SEQ ID NOS: 6 and 7 (GB AY310328)**

ggggtcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgagggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgcaactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctcttt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttggtgtgtttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaagaatgaccttcttgtgactcagctggccatcacag|atattaattggcgattcact
S R M T F F V T Q L A I T D I N W R F T
ggagacttcacggcacctgacctggttgccgagtgggtccgctatttgag|gttggtgctg
G D F T A P D L V C R V V R Y L Q V V L
ctctacgcctctacctacgtcctggtgtccctcagcatagacagataccatgccatcgtc
L Y A S T Y V L V S L S I D R Y H A I V
taccatgaagttccttcaaggag|aaaagcaagccagggtcctcattgtgatcgctgg
Y P M K F L Q G E K Q A R V L I V I A W
agcctgtcttttctgttctccattccaccctgatcatatttggaaggagacactgtcc
S L S F L F S I P T L I I F G K R T L S
aacgggtgaagtgcagtgctgggcccctgtggcctgacgactcctactggacccatacatg
N G E V Q C W A L W P D D S Y W T P Y M
accatcgtggccttcctggtgtacttcacccctctgacaatcatcag|catcatgtatggc
T I V A F L V Y F I P L T I I S I M Y G
attgtgatccgaactatttggtataaaagcaaacctacgaaacagtgatttccaactgc
I V I R T I W I K S K T Y E T V I S N C
tcag|atgggaaactgtgcagcagctataaccgaggactcatctcaaaggcaaaaatcaag
S D G K L C S S Y N R G L I S K A K I K
gctatcaagtatagcatcatcatcattcttg|ccttcactgtgctgttgagtgccatacttc
A I K Y S I I I I L A F I C C W S P Y F
ctggttgacatgttgacaatttcaacctccttcagacacccaggagcggtttctatgcc
L F D I L D N F N L L P D T Q E R F Y A
tctgtgatcattcagaacctgccagcattgaatagtgccatcaacccctcatctactgt
S V I I Q N L P A L N S A I N P L I Y C
gtcttcagcagctccatctcttccctgcag|ggatccgtctccgtcagctccaggag
V F S S S I S F P C R V I R L R Q L Q E
gctgcgctaagtctctgccctcaacgagagaactggaagggtacttgccaggtgtacct
A A L M L C P Q R E N W K G T W P G V P
tcctgggctctccaagggtgacagctctcaccctgtgctgcaggtggccctgtgcctggt
S W A L P R *
gccacttctcactgcttaccagggcacaaggacaccagtgggtccaaaatgggtcacag
caggatggcctgcatcagattcaccagggagggtataagaaggcagac

11/42

Figure 4C**C SEQ ID NOS: 8 and 9 (GB AY310329)**

```
ggggtcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgcccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgagggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctcttt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttgctttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaagaatgaccttctttgtgactcagctggccatcacag|tataacaagcccacctgctt
S R M T F F V T Q L A I T V *
gagctgggctgcagtgccagggtaaacatccaaggcaccagtgaaaaatacagagaagg
taaaaggagcaagagttctgaagatggaacctgggatgggggaaagtttcttcaatcttt
cctaccaacaagaactccaatttttactcctataaccgtagaagtagaggtaattagga
tcatccagcaaagcttagaggcaaatatccctggatgaggatgccacagcttattttca
ttatatcttctcgattacagtgtggtaatgcatgttgatggaactacatattctttcag
aatgaaaggatttagagggtgcaagaatatcagcttgaaatttaaagttttttcataaac
aataaacaatgataattgaaaattc
```

Figure 4D

D SEQ ID NOS: 10 and 11 (GB AY310330)

gggctcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgagggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgcaactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctcttt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttgtgctttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaagaatgccttcttggtagctcagctggccatcacag|gttgtgctgctctacgcctc
S R M T F F V T Q L A I T G C A A L R L
tacctacgtcctggtgtccctcagcatagacagataccatgccatcgtctacccccgaa
Y L R P G V P Q H R Q I P C H R L P H E
gttccttcaaggag|aaaagcaagccaggggtcctcattgtgatcgccctggagcctgtcttt
V P S R R K A S Q G P H C D R L E P V F
tctgttctccattcccaccctgatcatatttgggaaggagacactgtccaacgggtgaagt
S V L H S H P D H I W E E D T V Q R *
gcagtgctgggacctgtggcctgacgactcctactggacccatacatgaccatcgtggc
cttccctgggtgacttcatccctctgacaatcatcagcatcatgtatggcattgtgatccg
aactatttggattaaaagcaaacctacgaaacagtgatttccaactgctcagatgggaa
actgtgcagcagctataaccgaggactcatctcaaaggcaaaaatcaaggctatcaagta
tagcatcatcatcattcttgccttcatctgctgttggagtccatacttccctgtttgacat
tttggacaatttcaacctccttccagacaccaggagcggttctatgcctctgtgatcat
tcagaacctgccagcattgaatagtgccatcaacccctcatctactgtgtcttcagcag
ctccatctcttcccctgcaggagcaaagatcacaggattccagaatgacgttccggga
gagaactgagaggcatgagatgcagattctgtccaagccagaattcatctagaccctagg
gcagtgccagtgctaggctgagcaccatcagctctcccagggtccttgtcacctgcttggg
cacgtgcatggaacccgagccaacttcacccacccctcgctattacctgggagatgcaca
agacaaatgttctaatactgcatgcactgcttaagtattggccaacacgaactccccag
ttattcatgccagccaggaaggaaacgccttccctccccaccattcccagccctcctcc
cactggccagcacctgaaccagtgaaacacaggcattagtgggtccagggtcctgggttgg
agccagtgagtagac

Figure 4E**E SEQ ID NOS: 12 and 13 (GB AY310331)**

```
ggggtcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgcccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgaggtcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgcactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctctt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttgtgctttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaagaatgaccttcttgtgactcagctggccatcacag|attctttcacaggactggtc
S R M T F F V T Q L A I T D S F T G L V
aacatcttgacagatattaattggcgattcactggagacttcacggcacctgacctgggt
N I L T D I N W R F T G D F T A P D L V
tgccgagtggtccgctatttgcag|aaaagcaagccagggtcctcattgtgatcgctgga
C R V V R Y L Q K S K P G S S L *
gcctgtcttttctgttctccattcccacctgatcatatttgggaagaggacactgtcca
acggtgaagtgcagtgctgggcccctgtggcctgacgactcctactggaccccatatga
ccatcgtggccttcctgggtgacttcatccctctgacaatcatcagcatcatgtatggca
ttgtgatccgaactatttggattaaaagcaaacctacgaaacagtgatttccaactgct
cagatgggaaactgtgcagcagctataaccgaggactcatctcaaaggcaaaaatcaagg
ctatcaagtatagcatcatcatcattcttgccttcatctgctgttggagtccatacttcc
tgtttgacattttggacaatttcaacctccttccagacacccaggagcgtttctatgcct
ctgtgatcattcagaacctgccagcattgaatagtgccatcaacccctcatctactgtg
tcttcagcagctccatctcttccctgcagggagcaaagatcacaggattccagaatga
cgttcogggagagaactgagaggcatgagatgcagattctgtccaagccagaattcatct
agaccctagggcagtgccagtgctaggctgagcaccatcagctctcccagggtccttgtca
cctgcttgggcacgtgcatggaacccgagccaacttcaccccaccctcgtcattacctgg
gagatgcacaagacaaatgttctaatactgactgcagtgcttaagtattggccaacacg
aactccccagttattcatgccagccaggaaggaaacgccttcttccccaccattcccag
ccctccttcccactggccagcacctgaacccagtgaaacacaggcattagtgggtccagggt
cctggcttggagccagtgagtagac
```


14/42

Figure 4F**F SEQ ID NOS: 14 and 15 (GB AY310332)**

```
ggggtcag
ggagggctctgtgcctccgttcagcagagctgcagctgctgccagctctcaggaggcaa
gctggactccctcactcagctgcaggagcaaggacagtgaggctcaaccccgctgagcc
atgccagccaacttcacagagggcagcttcgattccagtgggaccgggcagacgctggat
M P A N F T E G S F D S S G T G Q T L D
tcttccccagtggttgactgaaacagtgacttttactgaagtgggtggaaggaaaggaa
S S P V A C T E T V T F T E V V E G K E
tgggggttccttctactactcctttaag|actgagcaattgataactctgtgggtcctctt
W G S F Y Y S F K T E Q L I T L W V L F
gtttttaccattgttgaaactccgttgctgtttttccacatggaggagaaagaagaag
V F T I V G N S V V L F S T W R R K K K
tcaagaatgaccttcttgtgactcagctggccatcacag|aaaagcaagccagggtcctc
S R M T F F V T Q L A I T E K Q A R V L
attgtgatcgccctggagcctgtctttctgttctccattcccaccctgatcatattggg
I V I A W S L S F L F S I P T L I I F G
aagaggacactgtccaacggtgaagtgcagtgctgggcccctgtggcctgacgactcctac
K R T L S N G E V Q C W A L W P D D S Y
tggaccccatatcatgaccatcgtggccttccctgggtgtacttcatccctctgacaatcatc
W T P Y M T I V A F L V Y F I P L T I I
ag|catcatgtatggcattgtgatccgaactatttggattaaaagcaaacctacgaaaca
S I M Y G I V I R T I W I K S K T Y E T
gtgatttccaactgctcag|atgggaaactgtgcagcagctataaccgaggactcatctca
V I S N C S D G K L C S S Y N R G L I S
aaggcaaaaatcaaggctatcaagtatagcatcatcatcattcttg|ccttcatctgctgt
K A K I K A I K Y S I I I I L A F I C C
tggagtccatacttccctgtttgacattttggacaatttcaacctcctccagacaccag
W S P Y F L F D I L D N F N L L P D T Q
gagcgtttctatgcctctgtgatcattcagaacctgccagcattgaatagtgccatcaac
E R F Y A S V I I Q N L P A L N S A I N
cccccatctactgtgtcttcagcagctccatctcttcccctgcag|ggagcaaagatca
P L I Y C V F S S S I S F P C R E Q R S
caggattccagaatgacgttccgggagagaactgagaggcatgagatgcagattctgtcc
Q D S R M T F R E R T E R H E M Q I L S
aagccagaattcatctagaccctagggcagtgccagtgctaggctgagcaccatcagctc
K P E F I *
tcccagggtccttgctcacctgcttgggcacgtgcatggaacccgagccaacttcacccac
cctcgtcattacctgggagatgcacaagacaaatgttctaatactgcatgcactgctta
agtattggccaacacgaactccccagttattcatgccagccaggaaggaaacgccttcct
tccccaccattcccagccctccttcccactggccagcacctgaacccagtgaaacacaggc
attagtggtccagggtcctggcttgagccagtgagtagac
```

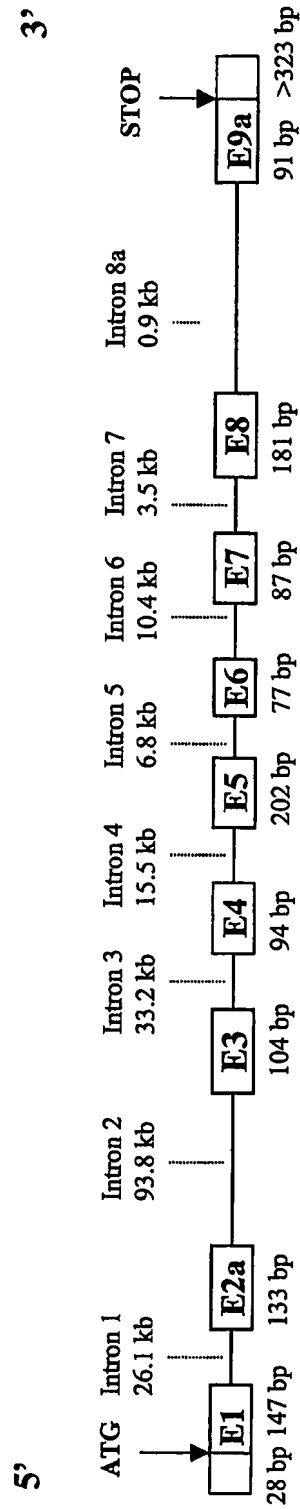


Figure 5A

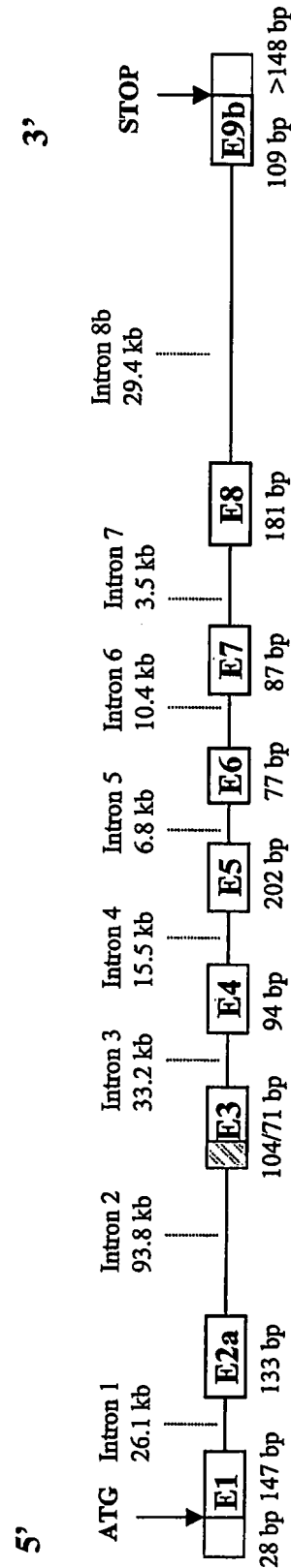


Figure 5B

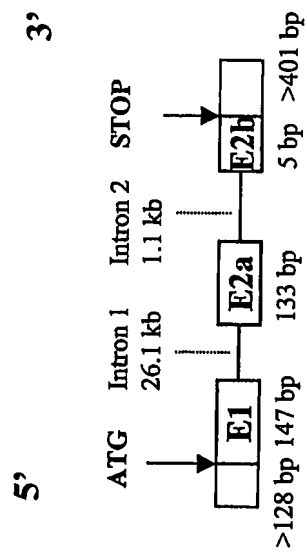


Figure 5C

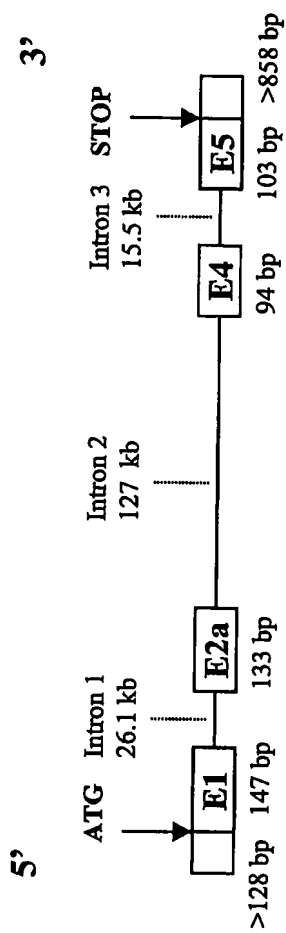


Figure 5D

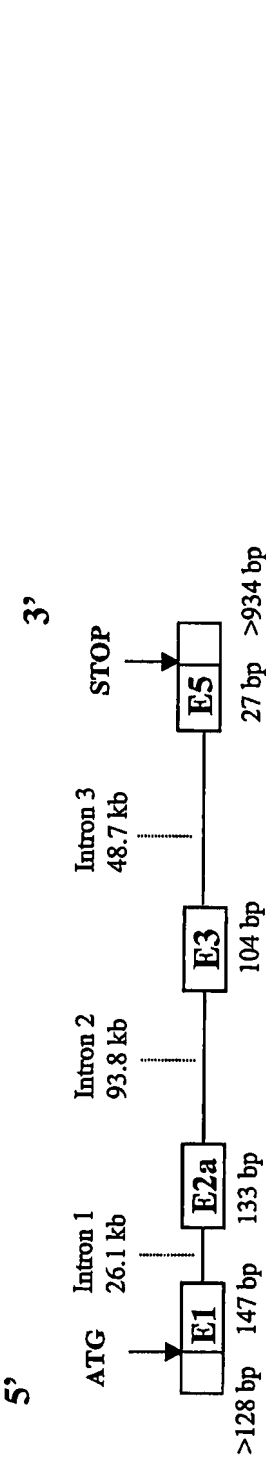


Figure 5E

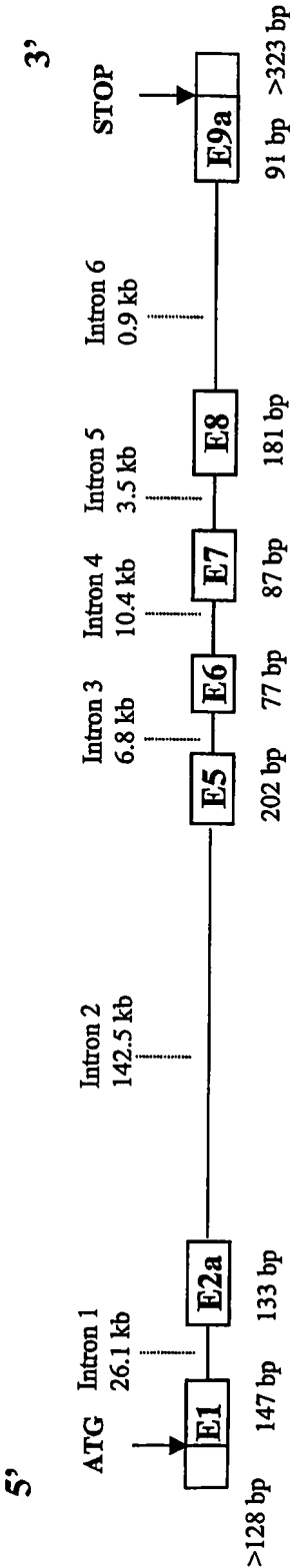


Figure 5F

18/42

MPANFTEGSFDSSGTGQTLDS SPVACTETVTFTEVVEGKEWGSFYYSFKT

TM1 CYTOLOOP1 TM2
EQLITLWVLFVFTIVGNSVVLFFSTWRRKKSRMTFFVTQLAITDSFTGLV

EXOLOOP1 TM3
NILTDINWRFTGDFTAPDLVCRVVRYLQVLLYASTYVLVSLSIDRYHAI

CYTOLOOP2 TM4 EXOLOOP2
VYPMKFLOGEKQARVLIVIAWSLSFLFSIPTLIIFGKRTLSNGEVQCWAL

TM5 CYTOLOOP3
WPDDSYWTPYMTIVAFLVYFIPLTIISIMYGIVIRTIWIKSKTYETVISN

TM6
CSDGKLCSSYNRGLISKAKIKAIKYSIIIIILAFICCWSPYFLFDILDNFN

EXOLOOP3 TM7
LLPDTQERFYASVIIQNLPALNSAINPLIYCVFSSSISFP CREQRSQDSR

MTFRERTERHEMQILSKPEFI*

Figure 6A

19/42

 B_{long}

MPANFTEGSFDSSGTGQTLDSSPVACTETVTFTEVVEGKEWGSFYYSFKT
 EQLITLWVLFVFTIVGNSVVLFSTM1WRRKKSRMTFFVTQLAITDSFTGLV
 NILTDINWRFTGDFTAPDLVCRVRYLQVLLYASTYVLVSLSIDRYHAI
 VYPMKFLQGEKQARVLIVIAWSLSFLFSIPTLIIFGKRTLSNGEVQCWAL
 WPDDSYWTPYMTIVAFVLVYFIPLTIISIMYGIVIRTIWIKSKTYETVISN
 CSDGKLCSSYNRGLISKAKIKAIKYSIIIIILAFICCWSPYFLFDILDNFN
 LLPDTQERFYASVIIQNLPALNSAINPLIYCVFSSSISFPCRVRILRQLQ
 EAALMLCPQRENWKGTPGVPSWALPR*

 B_{short} (33 bp deletion in exon 3)

MPANFTEGSFDSSGTGQTLDSSPVACTETVTFTEVVEGKEWGSF
 YYSFKTEQLITLWVLFVFTIVGNSVVLFSWRRKKSRMTFFVT
 QLAITDINWRFTGDFTAPDLVCRVRYLQVLLYASTYVLVSLSI
 DRYHAIVYPMKFLQGEKQARVLIVIAWSLSFLFSIPTLIIFGKRTLS
 NGEVQCWALWPDDSYWTPYMTIVAFVLVYFIPLTIISIMYGIVIRTI
 WIKSKTYETVISNCSDGKLCSSYNRGLISKAKIKAIKYSIIIIILAFIC
 CWSPYFLFDILDNFNLLPDTQERFYASVIIQNLPALNSAINPLIYC
 VFSSSISFPCRVRILRQLQEAALMLCPQRENWKGTPGVPSWALPR*

Figure 6B

20/42

C

MPANFTEGSFDSSGTGQTL DSSPVACTETVTFTEVVEGKEWGSFYYSFKT
EQLITLWVLFVFTIVGNSVVLFS TWRRKKKSRMTFFVTQLAITV*

D

MPANFTEGSFDSSGTGQTL DSSPVACTETVTFTEVVEGKEWGSFYYSFKT
EQLITLWVLFVFTIVGNSVVLFS TWRRKKKSRMTFFVTQLAITGCAALRL
YLRPGVPQHRQIPCHRLPHEVPSRRKASQGPHCDRLEPVFSVLHSHPDHI
WEEDTVQR*

Figure 6C and 6D

21/42

E

MPANFTEGSFDSSGTGQTLDSSPVACTETVTFTEVVEGKEWGSFYYSFKTEQLITLWVLF
VFTIVGNSVVLFFSTWRRKKKSRMTFFVTQLAITSFTGLVNILTDINWRFTGDFTAPDLVC
RVVRYLQKSKPGSSL*

F

MPANFTEGSFDSSGTGQTLDSSPVACTETVTFTEVVEGKEWGSFYYSFKTEQLITLWVLF
VFTIVGNSVVLFFSTWRRKKKSRMTFFVTQLAITEKQARVLIVIAWSLSFLFSIPTLIIFG
KRTLSNGEVQCWALWPDDSYWTPYMTIVAFLVYFIPLTIISIMYGIVIRTIWIKSKTYET
VISNCSDGKLCSSYNRGLISKAKIKAIKYSIIIIILAFICCWSPYFLFDILDNFNLLPDTQ
ERFYASVIIQNLPALNSAINPLIYCVFSSSISFPCREQRSQDSRMTFRERTERHEMQILS
KPEFI*

Figure 6E and 6F

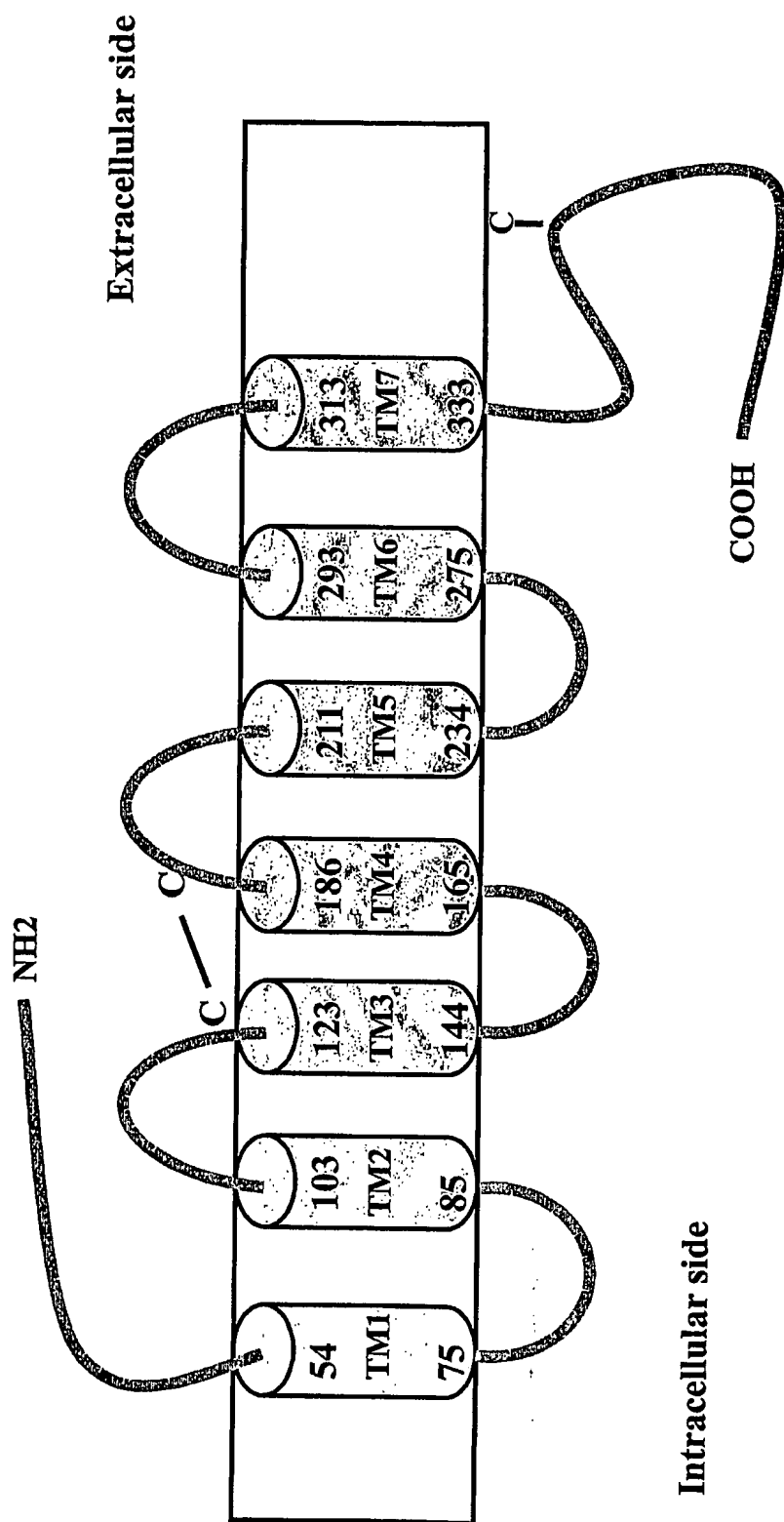


Fig. 7

23/42

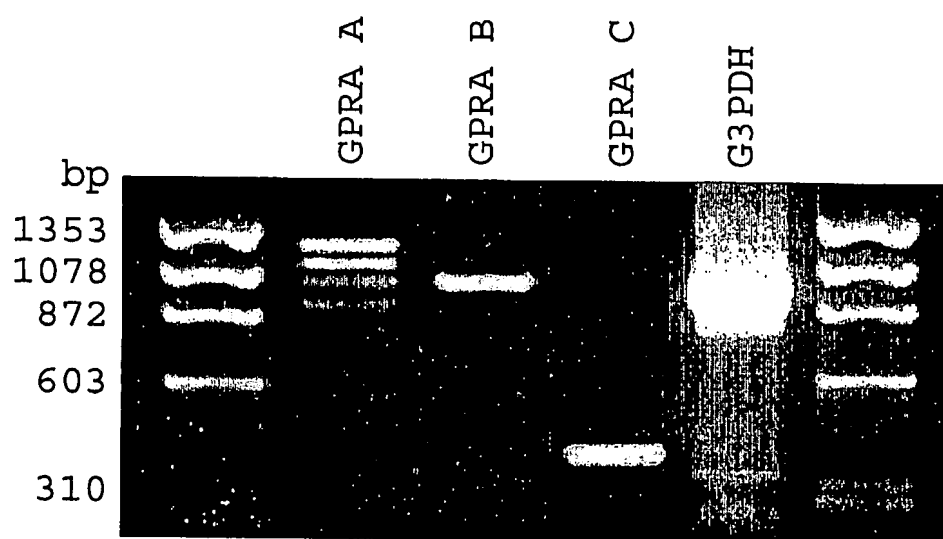


Figure 8

24/42

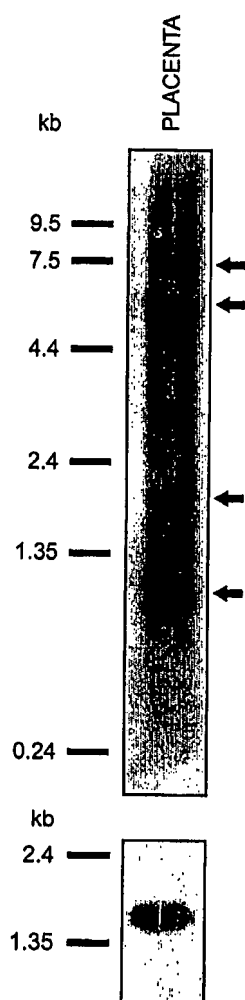


Figure 9

25/42

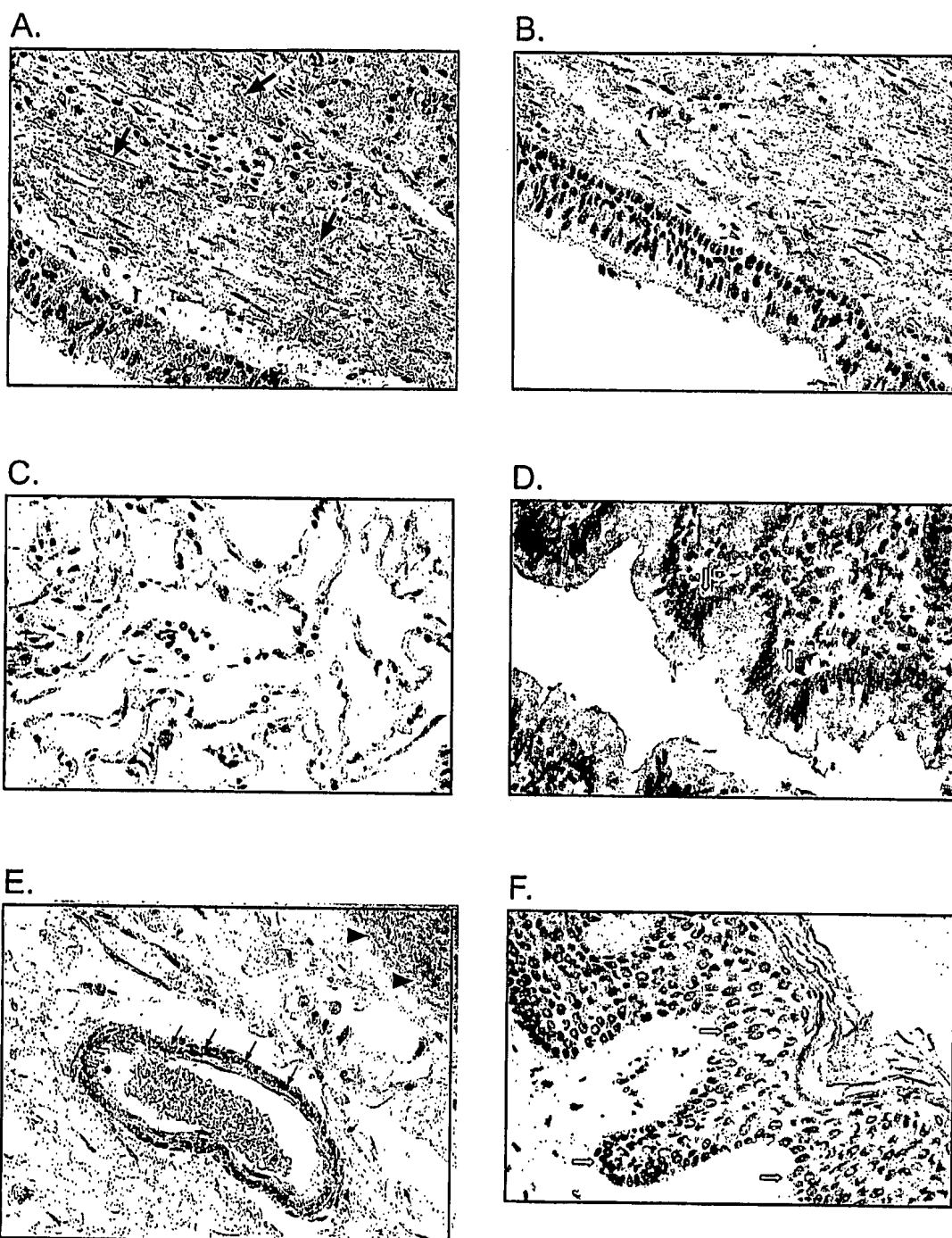


Fig. 10A, Fig. 10B, Fig. 10C, Fig. 10D, Fig. 10E, Fig. 10F

26/42

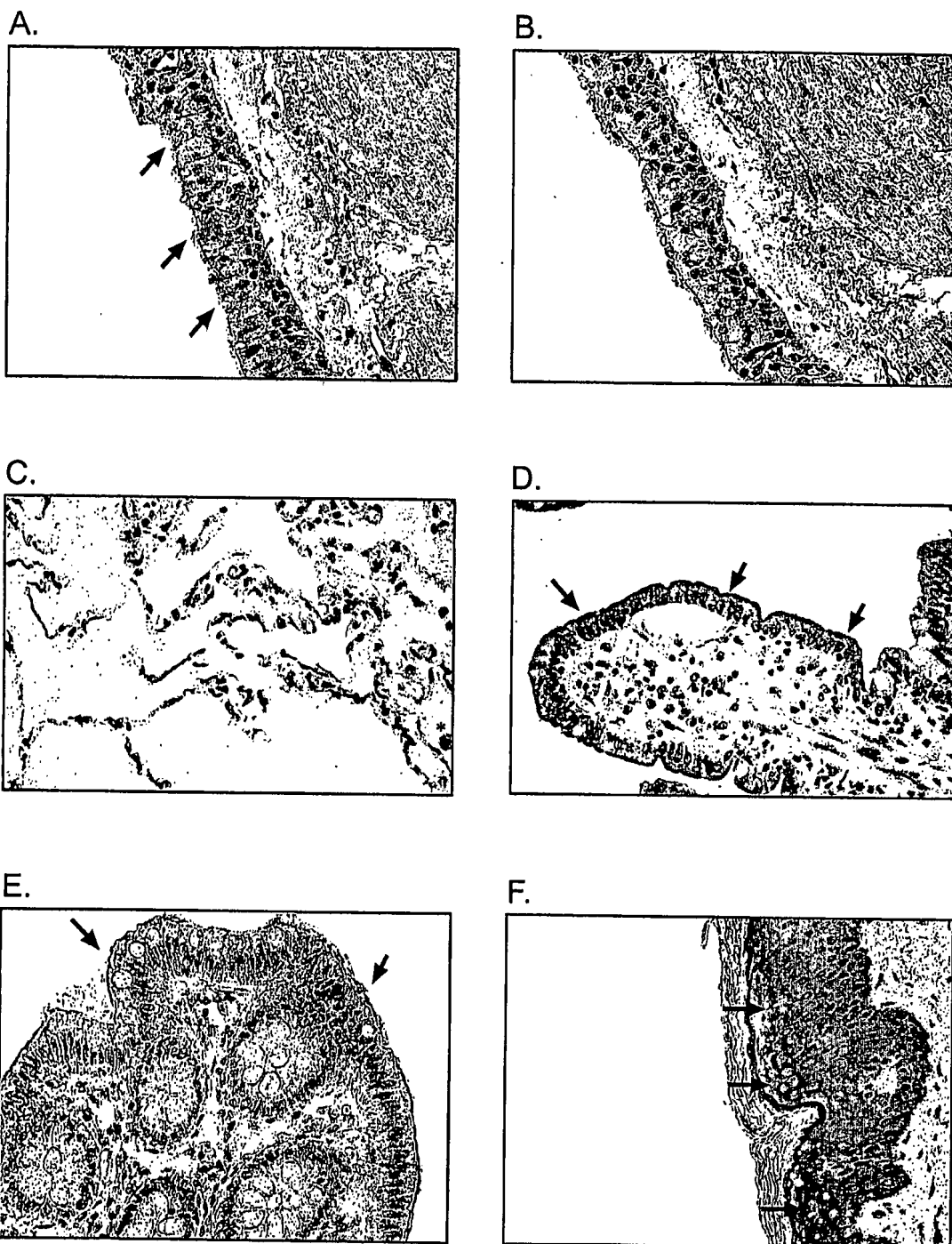


Fig. 11A, Fig. 11B, Fig. 11C, Fig. 11D, Fig. 11E, Fig. 11F

27/42

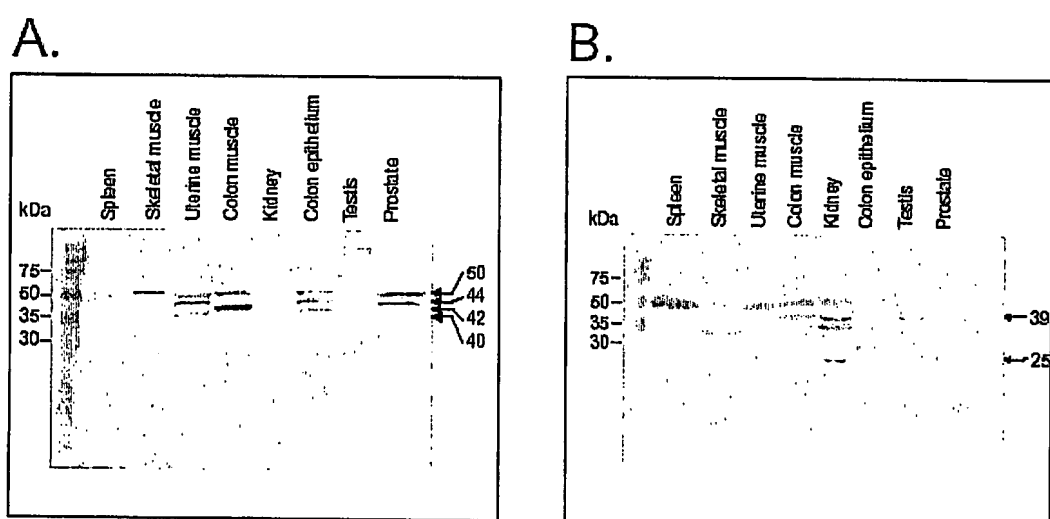


Fig. 12A and 12B

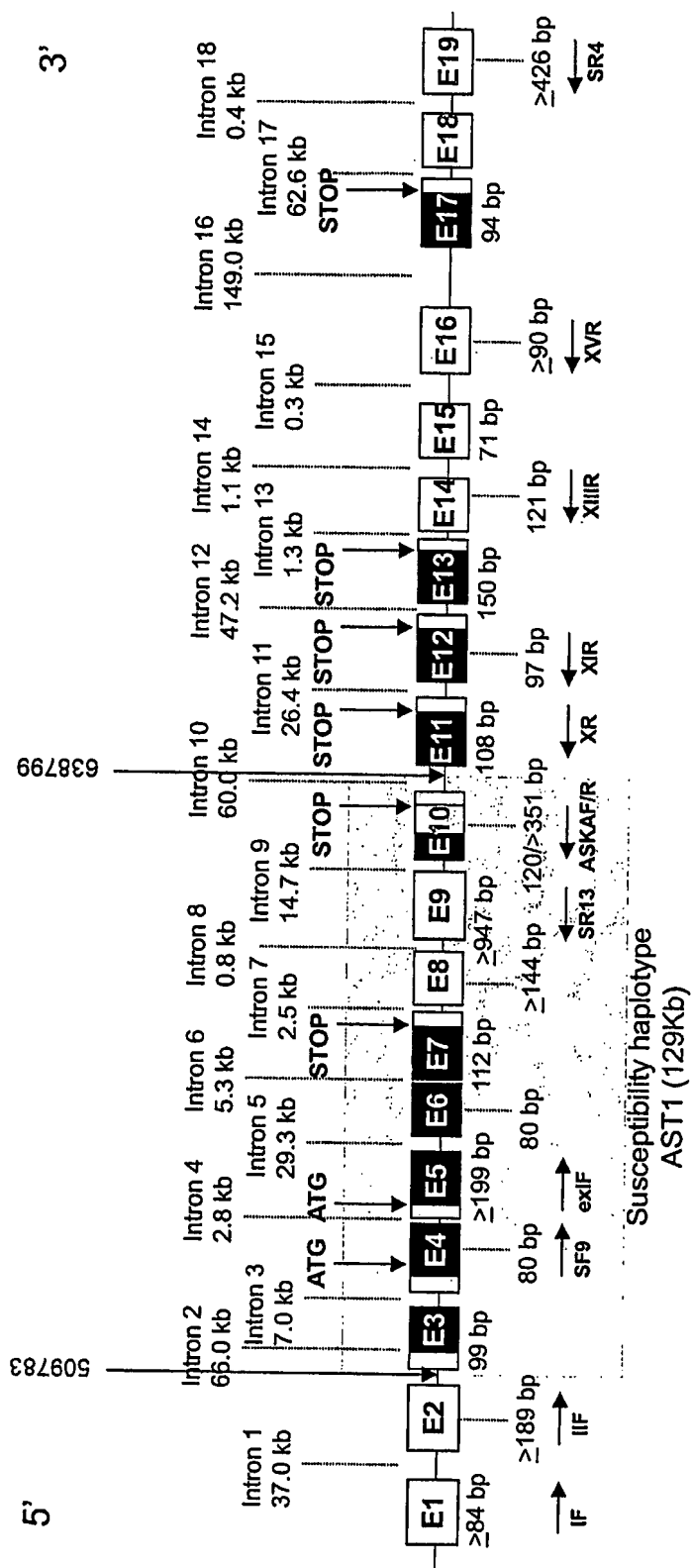


Fig. 13

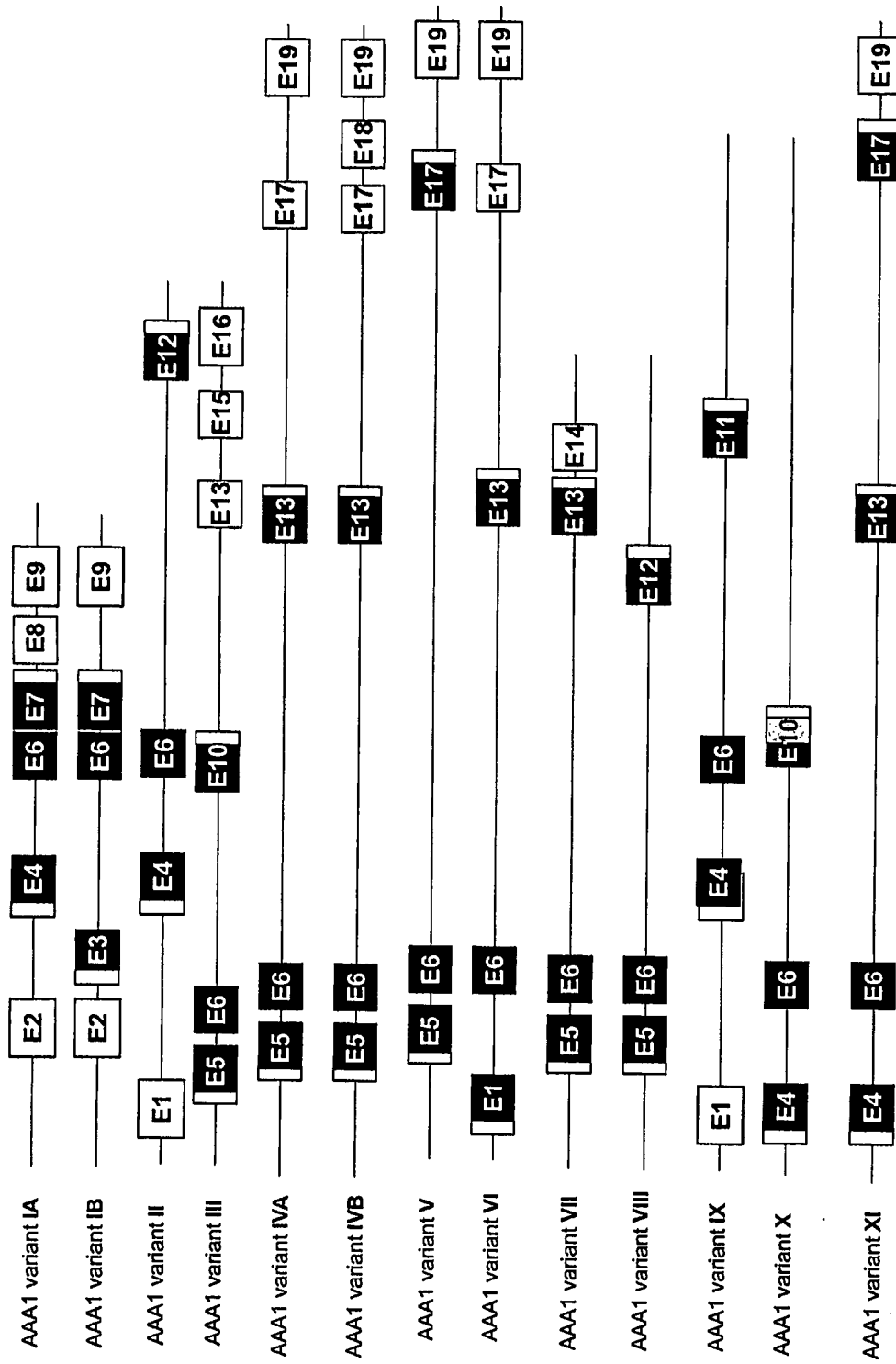


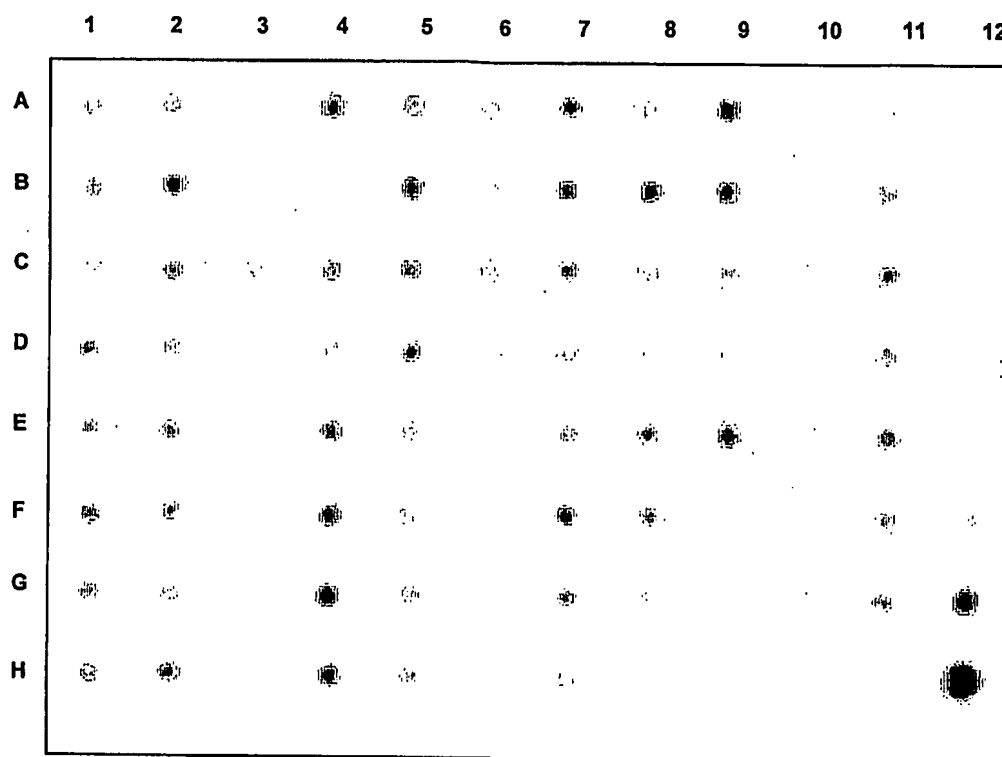
Fig. 14

I	MQCEAGAGAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQBELQSVRWYFPAAGLWVKDT
II	MQCEAGAGAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQVVDIHCSKQHVLLKRFVPSFNGLGTFPKN
III	MKKLKHRRPAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
IV	MKKLKHRRPAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
V	MKKLKHRRPAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
VI	MPLDLMLERLKTLGDYWKAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
VII	MKKLKHRRPAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
VIII	MKKLKHRRPAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
IX	MQCEAGAGAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
X	MQCEAGAGAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD
XI	MQCEAGAGAYVRRNAGRQFSHCNLSHCHQFLLVRRKKQENKD

*) shared protein coding sequence, different 3'-untranslated regions

Figure 15

31/42



	1	2	3	4	5	6	7	8	9	10	11	12
A	whole brain	cerebellum, left		heart	esophagus	colon, transverse	kidney	lung	liver	leukemia, HL-60	fetal brain	yeast total RNA
B	cerebral cortex	cerebellum, right	accumbens nucleus	norta	stomach	colon, descending	skeletal muscle	placenta	pancreas	HeLa S3	fetal heart	yeast tRNA
C	frontal lobe	corpus callosum	thalamus	atrium, left	duodenum	rectum	spleen	bladder	adrenal gland	leukemia, K-562	fetal kidney	E. coli rRNA
D	parietal lobe	amygdala		atrium, right	jejunum		thyroid	uterus	thyroid gland	leukemia, MOLT-4	fetal liver	E. coli DNA
E	occipital lobe	caudate nucleus	spinal cord	ventricle, left	ileum		peripheral blood leukocyte	prostate	salivary gland	Burkitt's lymphoma, Raji	fetal spleen	Poly (A)
F	temporal lobe	hippocampus		ventricle, right	ileocecum		lymph node	testis		Burkitt's lymphoma, Daudi	fetal thymus	human Cyt. J DNA
G	p. g.* of cerebral cortex	medulla oblongata		inter-ventricular septum	appendix		bone marrow	ovary		colorectal adenocarcinoma, SW620	fetal lung	human DNA 100 ng
H	pons	putamen		apex of the heart	colon, ascending		trachea			lung carcinoma, A549		human DNA 500 ng

* para-control gyrus

Figure 16

32/42

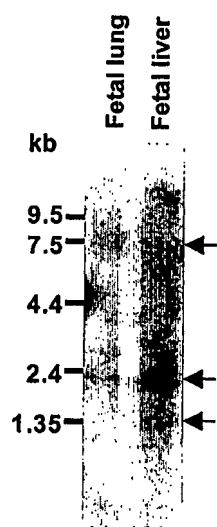


Figure 17

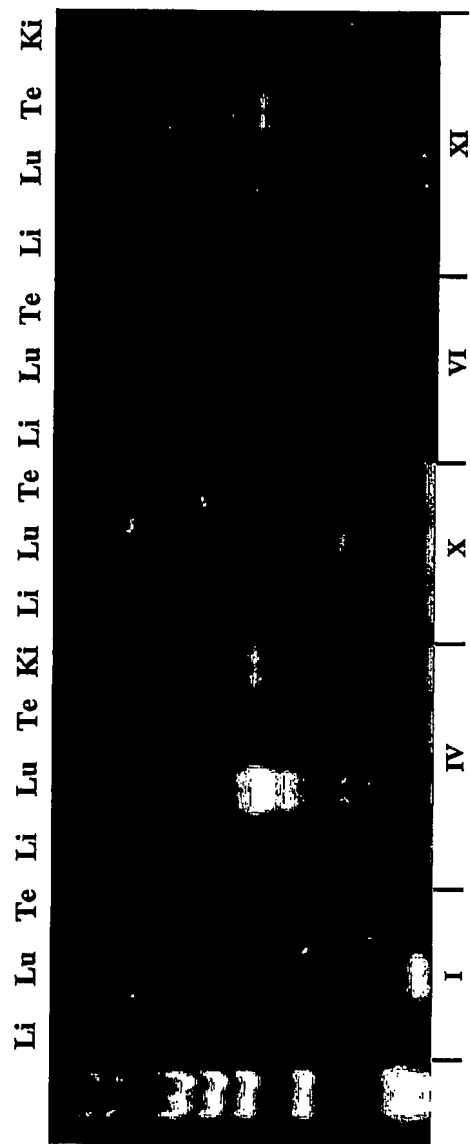


Figure 18

34/42

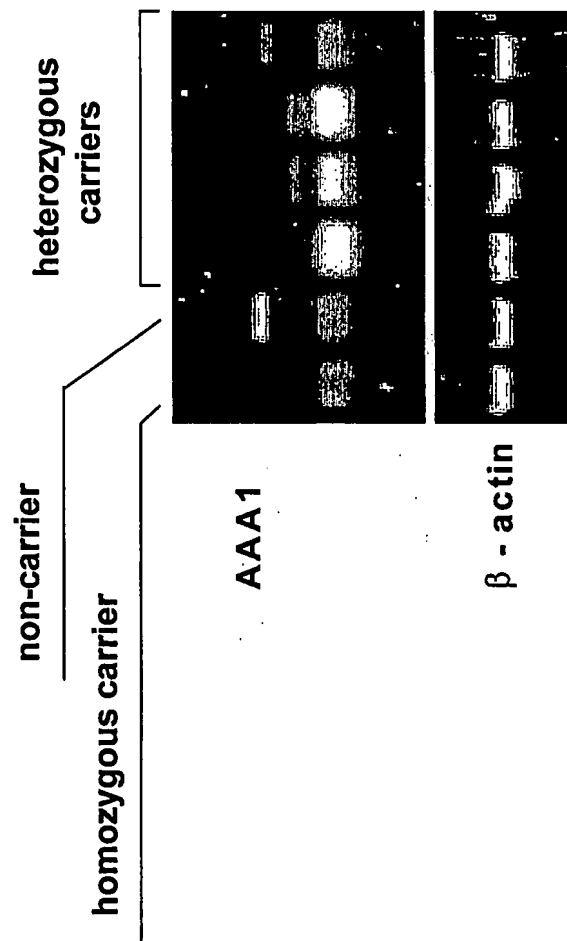


Figure 19

35/42

Figure 20

I exons 2, 4, 6, 7, 8, 9 GB AY312364 SEQ ID NOS: 16 and 17

gacacagagaagaccactcccacctccccgagtgcaaggtgtgaagggacagatcttttaac
catgcctgcccccttatacttgctgttcatagaattgcaactgaaagtgaccatgaggat
ccactggatggagttacttctttcttaagtgaggaggctaagatctggagtgacttctcc
ccagatctttgtatacctgactctgtttcagcatccgcttcccaagaatgcagtggtgaa
M Q C E
gcaggagcttatgtgagaagaaacgcagggagacagttcagtcactgcaatcttcatgcc
A G A Y V R R N A G R Q F S H C N L H A
catcagtttttgtgagaagaaaacaagagctccaatctgttagatgggtattttgaagca
H Q F L V R R K Q E L Q S V R W Y F E A
gggtctttgggtaaaggacacctagaccagtggaaggtcatgggtgattattattggacaat
G L W V K D T
gggacatcactctgctattttgaacaaataagactttttcctgacatgcatctggaggcag
gtcaaagctccaggccaactccaagtttctgatggggtctctagccaatggaaggcttct
tctccttcaattgctgactcttcaggactcttaataactgcaaagtgagaaaatgagac
aggttgcaactgagggctgttagccagacagagctctcgaactggaagtccatctagatgtt
ttgcataagagaatggaaacaatctgtctgtgatttagggacatactctggcagcaatat
gggaatacagtttcaatcctcatatacaaaacaggtatgaaatacatatttttagtaa
gggtgccagctgtatgaaaaatccatttcttatttcccataatgtttctgaaatgtcttag
cagtgcatagagacagcatgtcatcttttctagggactgtgtgttattgcatttttct
aggggaagatcttttctaggtcacctgctccttcgctaaagctctgaccaatctagcttgc
taacctgtgactccattttcctaagtcctgagagagaaaaacgctttgcagcaaattatg
ccaggcatccttgtgtctaaatgaaaaaaggaaaaagcctccttccctccctctgttgag
aagtgcacgggtccacatatgcatgcacagcatatactgtgaggtatttgcagtccttgg
gggtgctttgataactggccaggttgctgttctatttttccacattctattaatcctcct
acaggcagttattaggtattgagtgtcacacacccctggcatagtcaccacatgccatt
agctccagataaacttccagaaaaagtcacccccacttctctcagctgcctgccaac
gctggacaccttctcaccaagccagcaggacagagaaaagcctgggctttaagatcaaac
aaacacagcttcaaattaggactctgtcacttcctgtgtactgggactttgctgagtat
gtgggttctcatctgtaaaacagagaaagatgattatctcccaatctttctatgttatat
gtttgaattaaataaggtactctccatgaa

37/42

Figure 20 (continued)

III exons 5, 6, 10a, 13, 15, 16

GB AY 312367 SEQ ID NOS: 22 and 23

tctaggactcagaaatatagatgttagtaagagcaaacagacataacagataaacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaaccctatggggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccacagtttcttgtgagaagaaaacaagaaaacaaggactgaaatcca
N L H A H Q F L V R R K Q E N K D
cacaggaaggtggcagtgaaactccacagacggacctggacgcctcaacactcctggcctt
acctcccttgctgaacgtctcaagttctctgcgttcaggactggcaacgcctgcttcct
cctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcagtgtgcag
gagcactgaggcatcccaggtgtgacactcttccacctagagcattccgtctctcatcct
ctgccatgtgacgtgggcttctttaaacaattaatcccaagtgaagacattttatttct
tctgtacctaatgacctgagcaatccttctctgctgaacctggtagtgcattcttagaa
gtgaagacacaattaaacacatggtcatttcttcattatcgttggtact

IV exons 5, 6, 13, 17, 19

GB AY312368 SEQ ID NOS: 24 and 25

tctaggactcagaaatatagatgttagtaagagcaaacagacataacagataaacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaaccctatggggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccacagtttcttgtgagaagaaaacaagactggcaacgcctgcttcct
N L H A H Q F L V R R K Q D W Q R L L P
cctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcagtgtgcag
P L S C Q V G S P G C S A R K R S H V Q
gagcactgaggcatcccaggtgtgacactcttccacctagagcattccgtctctcatcct
E H
ctgccatgtagcaaaactgctatgcatccttcagctgcaagggattgaatgctatcaacaa
ccatacaagtggaagcagatgcttccctagctgagcctcaggctttttgatggaattg
ctacaacttggtgcatgctgctcctaaaagaaatactcaggaattgtctcataaagtcc
tcacctactggcaaaaacaagatgttctactcccaggttgactttttcaagccccaagat
gttgagtcagccatttctccaaggatctcgatttcttttaattggaataaacattaaaca
ccaaatataagcctcgctgtccacatgcgtattggggacaagatgaaacctgcttcag
gctactttggcagcagaactgaaaaggcttttttccagatatatgatttctcatcgac
agggttgcacagccctctttattgttcgtgtaaatgacacccttggtatctgaacaatata
caccaggacaattgtgtgcaacagttctacaaactgatatttctaatta

38/42

Figure 20 (continued)

IVB exons: 5, 6, 13, 17, 18, 19

SEQ ID NOS: 26 and 27

gtctaggactcagaaatatagatgtagtaagagcaaacagacataacagataaacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaacctatgggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccacagtttcttgtagagaagaaacaagactggcaacgcctgcttcct
N L H A H Q F L V R R K Q D W Q R L L P
cctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcagtgag
P L S C Q V G S P G C S A R K R S H V Q
gagcactgaggcatcccagggtgtgacactcttcacctagagcattccgtctctcactc
E H
ctgccatgtagcaaactgctatgcatccttcagctgcaagggattgaatgctatcaacaa
ccatacaagtgtagaagcagatgcttccctagctgagcctcagctctgttttcctcagtc
ctaggggcagatttctaatttggtacagaaaaaaaagtactcctgccaattcccactctg
gaatgactgaagaaaagaaaatgcatgtgattatttttaaaggctttttgatggaattgct
acaacttggtgcatgcctgctcctaaaagaaatactcaggaattgtctcataaagtcctc
acctactggcaaaaacaagatgttctactcccaggttgactttttcaagcccaagatgt
tgagtcagccatttctccaaggatctcgatttccttttaaggaaaataacattaaacacc
aaatataagcctcgctgtcccacatgcgtattggggacaagatgaaacctgcttccaggc
tactttggcagcagaactgaaaaaggcttttttccagatatatgatttctcatcgacag
ggttgcacagccctctttattgttcgtgtaaatgacacccttgatctgaacaatacaca
ccaggacaattgtgtgcaacagttctacaaactgatatttctaatta

V exons 5, 6, 17, 19

GB AY 312369 SEQ ID NOS: 28 and 29

tctaggactcagaaatatagatgtagtaagagcaaacagacataacagataaacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaacctatgggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccacagtttcttgtagagaagaaacaagcaaaactgctatgcatcctt
N L H A H Q F L V R R K Q S K L L C I L
cagctgcaagggattgaatgctatcaacaaccatacaagtggagaagcagatgcttccct
Q L Q G I E C Y Q Q P Y K W R S R C F P
agctgagcctcaggctttttgatggaattgctacaacttggtgcatgcctgctcctaaaa
S
gaaatactcaggaattgtctcataaagtcctcacctactggcaaaaacaagatgttctac
tcccagggtgactttttcaagcccaagatgttgagtcagccatttctccaaggatctcga
tttcttttaaggaaaataacattaaacaccaaataaagcctcgctgtcccacatgcg
tattggggacaagatgaaacctgcttccaggctactttggcagcagaactgaaaaaggct
tttttccagatatatgatttctcatcgacagggttgacagccctctttattgttcgtg
taaataacaccccttgatctgaacaatacacaccaggacaattgtgtgcaacagttctac
aaactgatatttctaatta

36/42

Figure 20 (continued)

IB: exons 2, 3, 6, 7, 9

GB AY312365 SEQ ID NOS: 18 and 19

gacacagagaagaccactcccacctccccgagtgcaaggtgtgaaggacagatctttta
 accatgcctgcccccttatacttgctgttcatagaattgcaactgaaagtgaccatgagg
 atccactggatggagttacttctttcttaagtgaggaggctaagatctgagttcttcaca
 tctctctgtagataaaaatttccggtctgggttcacattcctctgtcagaagaactttctt
 taatgtttcttaaagtacaggtctgctgcttatgtgagaagaaacgcaggagacagttc
 C F L K Y R S A A Y V R R N A G R Q F
 agtcactgcaatcttcatgcccacagtttcttgtgagaagaaaacaagagctccaatct
 S H C N L H A H Q F L V R R K Q E L Q S
 gttagatgggtattttgaagcaggtcttgggtaaaggacacctagaccagtgagggtca
 V R W Y F E A G L W V K D T
 tgggtgattattattggacaatgggacatcactctgctattaagtgagaaaatgagacagg
 ttgactgagggctgttagccagacagagtctcgaactggaagtccatctagatgttttg
 cataagagaatggaaacaatctgtctgtgatttagggacatactctggcagcaatatggg
 aatacagtttcaatcctcattaacaaaacaggtatgaaatacatatttatttagtaagg
 gccagctgtatgaaaaatccatttcttatttcccataatgtttctgaaatgtcttagcag
 tgcataagacagcatgtcatcatttcttagggactgtgtgttattgcatttttcttagg
 gaagatcttttctaggtcacctgctccttcgctaaagctctgaccaatctagcttgctaa
 cctgtgactccatttccctaagtcctgagagagaaaaacgctttgcagcaaatatgcc
 ggcatccttgtgtctaaatgaaaaaaggaaaaagcctccttccctcctctgttgagaag
 tgcacgggtccacatatgcacagcatatactgtgaggggtatttgcagtcacctgggt
 tgctttgataactggccaggttgctgttctatttttccacattctattaatcctcctaca
 gccagttattaggtattgagtgctcacacaccctggcatagtcaccacatgccattagc
 tccagataaaacttccagaaaaaagtcacatccccacttctctcagctgcctgccaacgct
 ggacaccttctcaccaagccagcaggacagagaaaagcctgggctttaagatcaaaacaaa
 cacagcttcaaattaggactctgtcacttctgtgtactgggcactttgtgagtatgtg
 gtttctcatctgtaaaacagagaaagatgattatctcccaatctttctatgttatatgtt
 tgaattaaataaggtaactctccatgaa

II exons 1, 4, 6, 12

GB AY312366 SEQ ID NOS: 20 and 21

gctgatgggtggaaggagaatgagtcctgatgcctttggacttgatgctggaagacttaa
 gactttgggggactactggaaggagtgacttctccccagattttgtatacctgactct
 gtttcagcatccgcttcccaaagaatgcagtggtgaagcaggagcttatgtgagaagaac
 M Q C E A G A Y V R R N
 gcaggagacagttcagtcactgcaatcttcatgcccacagtttcttgtgagaagaaaa
 A G R Q F S H C N L H A H Q F L V R R K
 caagtggatatacactgttccaagcagcatgtgttgaaaagatttgtcttttccccattt
 Q V D I H C S K Q H V L K R F V F S P F
 aatggctcttggtaccttttctcaaaaattgaccatatatga
 N G L G T F L K N

39/42

Figure 20 (continued)

VI exons 1, 6, 13, 17, 19

GB AY 312370 SEQ ID NOS: 30 and 31

gctgatggaggagaatgagtcctctgatgcctttggacttgatgctggaaagacttaag
M P L D L M L E R L K
actttgggggactactggaaagcttatgtgagaagaaacgcagggagacagttcagtcac
T L G D Y W K A Y V R R N A G R Q F S H
tgcaatcttcatgcccacagtttcttgtagagaagaaaacaagactggcaacgcctgctt
C N L H A H Q F L V R R K Q D W Q R L L
cctcctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcattgtg
P P L S C Q V G S P G C S A R K R S H V
caggagcactgaggcatcccaggtgtgacactcttccacctagagcattccgtctctcat
Q E H
cctctgccatgtagcaaaactgctatgcattcagctgcaagggattgaatgctatcaa
caaccatacaagtggagaagcagatgcttccctagctgagcctcaggcctttttgatggaa
ttgtacaacttggatgcatgcctgctcctaaaagaaataactcaggaattgtctcataaag
tcctcacctactggcaaaaacaagatgttctactcccagggtgactttttcaagcccaa
gatgttgagtcagccattctccaaggatctcgtatttcccttttaattggaaaaataacattaa
acaccaaataaagcctcgtctgtccacatgcgtattggggacaagatgaaacctgcttc
caggctactttggcagcagaactgaaaaaggccttttttccagatatatgattttctcatc
gacaggggttgacagccctctttattgttcgtgtaaatgacacccttggatctgaacaat
acacaccaggacaattgtgtgcaacagttctacaaactgatattttctaatta

VII exons 5, 6, 13, 14

GB AY 312371 SEQ ID NOS: 32 and 33

tctaggactcagaaatatagatgttagtaagagcaaacagacataacagataacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaccctatggggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccacagtttcttgtagagaagaaaacaagactggcaacgcctgcttctt
N L H A H Q F L V R R K Q D W Q R L L P
cctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcattgtgcag
P L S C Q V G S P G C S A R K R S H V Q
gagcactgaggcatcccaggtgtgacactcttccacctagagcattccgtctctcatcct
E H
ctgccatgtgccatgttttgaaccactagattagaggggtcaagcaatttcttggaaatttt
actctgaattctacgttagaccattttcatgtgtatacctcctctgagtcaccctcaggta
gggacatttt

Figure 20 (continued)

VIII exons 5,6,12

GB AY 312372 SEQ ID NOS: 34 and 35

tctaggactcagaaatatagatggttagtaagagcaaacagacataacagataaacacatac
aaagtgcctaccacatgctaaccactgctgcaggcactttctatagaagaactaatttaa
tcatcaccataaccctatggggtagatgatatttttacaacctccattttacagatgaag
M K
aaactgaagcatagacctgcttatgtgagaagaaacgcagggagacagttcagtcactgc
K L K H R P A Y V R R N A G R Q F S H C
aatcttcatgcccatcagtttcttgtgagaagaaaacaagtggatatacactgttccaag
N L H A H Q F L V R R K Q V D I H C S K
cagcatgtgtgaaaagatttgtcttttccccatttaattggtcttggtagcttttctcaaa
Q H V L K R F V F S P F N G L G T F L K
aattgaccatatatga
N

IX exons 1, 4, 6, 11

GB AY 312373 SEQ ID NOS: 36 and 37

gctgatgggtggaaggagaatgagtcctctgatgcctttggacttgatgctggaaagacttaa
gactttgggggactactggaaaggagtgacttctccccagatttttgtataacctgactct
gtttcagcatccgcttcccaaagaatgcagtggtgaagcaggagcttatgtgagaagaaa
M Q C E A G A Y V R R N
gcagggagacagttcagtcactgcaatcttcatgcccatcagtttcttgtgagaagaaa
A G R Q F S H C N L H A H Q F L V R R K
caagtttaggaaaacttctacaccttctttgttgggatgttctctggactaatgactcc
Q V
aggcgagaccaccgttgatcatgaactcactttgaaacagaagctgggttggttaagactg
gagctact

X exons 4, 6, 10b

GB AY 321515 SEQ ID NOS: 38 and 39

gagtgacttctccccagatttttgtataacctgactctgtttcagcatccgcttcccaaaga
atgcagtggtgaagcaggagcttatgtgagaagaaacgcagggagacagttcagtcactgc
M Q C E A G A Y V R R N A G R Q F S H C
aatcttcatgcccatcagtttcttgtgagaagaaaacaagaaaacaaggactgaaatcca
N L H A H Q F L V R R K Q E N K D
cacaggaaggtggcagtgaaactccacagacggacctggacgcctcaacactcctggcctt
acctcccttgctgaacgtctcaagtttctctgcgttcaggtaatgtataggaggggttatg
agggcagagaattcctaagctcattagtaaattgctcttcagaaaaagtgtttgaagcaa
agctaatttcccttcccaatatgagaagatttggccctaccagaaaaaggaaatgatttg
aatgtgcgcccaaaaaatatgttttctttcttttcttgggtgaacactcatcgaggagttac
tcttattagttccgcattttttattgccattt

41/42

Figure 20 (continued)

XI exons 4, 6, 13, 17, 19

GB AY 321516 SEQ ID NOS: 40 and 41

gagtgacttctccccagatTTTTgtatacctgactctgtttcagcatccgcttcccaaaga
atgcagtgtgaagcaggagcttatgtgagaagaaacgcagggagacagttcagtcactgc
M Q C E A G A Y V R R N A G R Q F S H C
aatcttcatgcccatcagtttcttgtgagaagaaaacaagactggcaacgcctgcttct
N L H A H Q F L V R R K Q D W Q R L L P
cctctgagctgtcaagtaggaagtccgggctgctctgctagaaagagaagtcagtgtgcag
P L S C Q V G S P G C S A R K R S H V Q
gagcactgaggcatcccggtgtgacactcttcacctagagcattccgtctctcatcct
E H
ctgccatgtagcaaactgctatgcacaccttcagctgcaagggattgaatgctatcaacaa
ccatacaagtggagaagcagatgcttccttagctgagcctcaggctttttgatggaattg
ctacaacttgggtgcatgcctgctcctaaaagaaatactcaggaattgtctcataaagtcc
tcacctactggcaaaaacaagatgttctactcccaggttgactttttcaagccccaagat
gttgagtcagccattctccaaggatctcgatttccttttaatggaaaataacattaaaca
ccaaatataagcctcgctgtccacatgcgtattggggacaagatgaaacctgcttcag
gctactttggcagcagaactgaaaaaggcttttttccagatatatgatttctcatcgac
agggttgacagccctctttattgttcgtgtaaatacaccccttgatctgaacaataca
caccaggacaattgtgtgcaacagttctacaaactgatatttctaatta

42/42

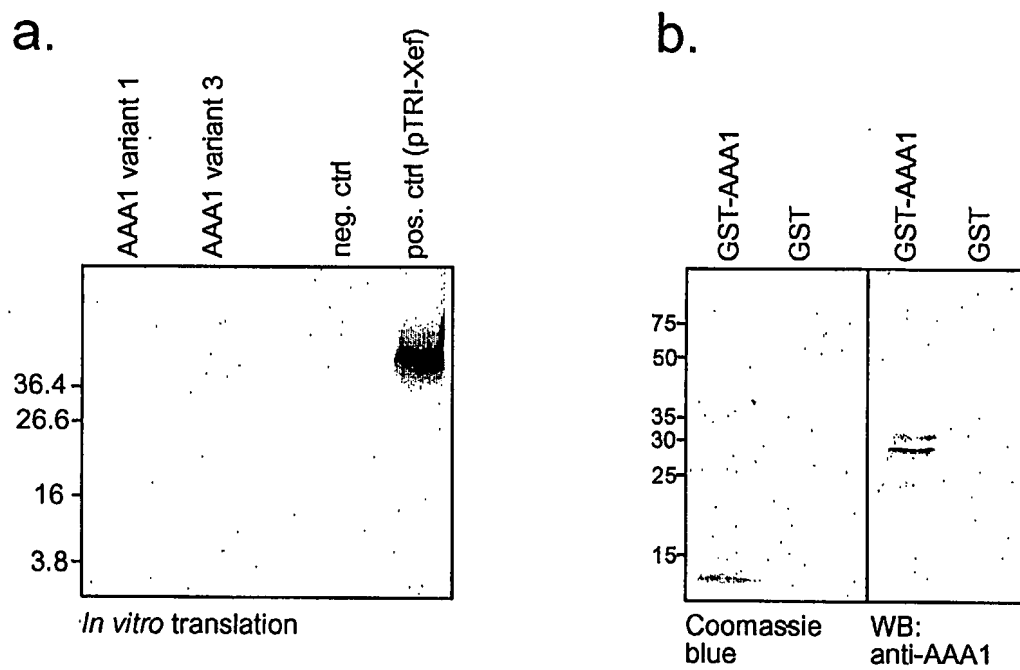


Fig. 21A and 21B

SEQUENCE LISTING

<110> Geneos Oy

<120> Asthma susceptibility locus

<130> 40564

<150> US 60/435,846

<151> 2002-12-20

<150> US 60/437,895

<151> 2003-01-03

<150> US 60/458,767

<151> 2003-03-26

<150> US 60/486,000

<151> 2003-07-09

<160> 41

<170> PatentIn version 3.0

<210> 1

<211> 129017

<212> DNA

<213> Homo sapiens

<400> 1

cggcaaatta aatgtagcta aagaagaaat cattgaactg gaatataggt ccaaagaaaa 60

taattagatt gaagcatgaa aaaagaaagg acaaaaaata caaaaagaaa ataaaaagca 120

tatggaagta gtgaaaagga ctcccagaaa aagataaagc ccagactgga tagaaacaac	180
acttatagtg gccataaatt ttcccaact ctttaaagac atcaaaacac aaatacataa	240
agcaccataa aggacaacag caagcaaaca aacaaatata aagagaatta aaacacatga	300
aatgataaga ctaagtaaaa gcccataat cttgaatttg aaatgaaaat atttatgcta	360
gctcatgatg tattttatgc ttcaggagtt ttttaaggata catacaaaaa tgtttatagg	420
taaactgata acatgtcttg gatttacttc aaaataattt gaggctgggg gtgggttgag	480
atgtatagat gtacaaaatt gatattgata actgttgagg gttggtgggt agtacattgg	540
attctttatt ttttttatac tattctctct acttttatgt aagttaaaga ttttttcat	600
aataaaattc aaaacaaaat tatgcatgat tataatgaaa caaaataaag agatctgggt	660
agttaaagat gttaaagtcc ttggattgcc tggatgtagt aaaagtacta atttctatta	720
gattttaatg tcaaggaggt gtgttaaaat ctctaaagta atcctcccaa attatgaaaa	780
ctgacattac tagcaagcta aagaaggat aaaatgcagt atttgaatat gcaaatatat	840
atgtttatat atagagttat tcccagtat ccatggagga ttagttccag aatcccctgc	900
caatacaaaa attcgaggat gcttaagtcc cttacataaa atagatagta tttgcatata	960
acctacatac accctcctgt ataactttaag tcatctctaa attacttata atttctaattg	1020
caatgtaaatt tctatataaa aatttttata ctatatctt aaaatttgta tcattttatt	1080
attgtattgt tattttatac tttcttaata tttttaatct gcagttgata gaatccatgg	1140
atgtggaacc cacagataaa aaggccaat tataatacat acatattata taaaatcaat	1200
ctaagttaag acaggagaga aaaaatatta gaattattag aacaagtaag aagaataggc	1260
cagtacactt aacctcaaatt gtatcagtaa ttgcattaat tgtaaatgga ctaaattatc	1320
ccattaaaat aacaagattg tcagagtaga tttttaaatt acctatttta atttgtttat	1380
aaaagggtga caccataaat aaaatgatgc aggttaaaat taaaaggaaa aggatggtaa	1440
agtaaaatga taaaaataa tgaaacaaaa caaacacaa agtaaaagta agccaatgca	1500
gctacatatg tgtatcaaaa gtttataaat ttgctgtgat tcttggaatc acttgtagct	1560
ttatctgtgt cttagaattt gctttgaaag gaatacaaag acaggagggc ttctaaagct	1620
agtaatgttc tatttcttga tcttgcttgt agttacgtgg gcgtgttctt cttgtaataa	1680
ttcaccaaat tatatactta tgattttgta tacttttctt tatgtgtttt atttcaatac	1740
aaagtttatt taaaaatat atcacattaa tccagggcag ggggttgcaat ctcaaatggc	1800
tttaggaatc aagcaggcaa tgaatgtaaa catgtaaaag agccaagggg aagagacaga	1860
atgaaaacag agataaaaac cagagaatga atgccctgcc taaagacatt aaaatcaatt	1920

tttgtcaaac actgtgctga tcaaacaaat cacacatgca ggccagattt gacccaaggg	1980
tgacttttctt acaacaactg atatataaaa tttaaacatc agaattttgaa aagtgtgcc	2040
ctataaaatt ctagcccacc tctgaagtat gaaccacact tcaattttgtg aggtagaata	2100
ttaatcttta ctggtggatt gtgagaagtg aggccaagaa gttggggagg ttaacattta	2160
caatggttta gacaaatcat atgtttgaca aggtttaaaa tcccttctgt tcaaagaata	2220
tcttatatta ggattttctt ctctaattgt gtttattatt gtttttatta gtataatttt	2280
taaagaaatc tgaactattg cacgcctcag agaccataac agtgagaaca aaaatataaa	2340
ggaaaatttt caaatgatgg aaagtgtaac ccatataacc cgaaatcttc atgattttca	2400
gatgaccatc tgattaagga gaaaagtcaa gataagcaca tattcatatc tgcttatttc	2460
tagaattggt ctatcagaag accatgaatt ggtcatatga ggggaagaaa cactgacatg	2520
atactggctc tattaggatg atattaaaag aggctgctgg ctcaagctta gcatggatga	2580
gttgttctca cataaagatg ctgttattga ccagggtgtg tggtcacgc ctataatccc	2640
agcactttgg gaggctttgg tgagtagatc ccctgtgtcc aggagttcaa gaccagcctg	2700
agcaacatag tgagaccctg tctctactaa aaaatttttt aaattaaaag ttaaaaaatt	2760
aatgatggt gctattgtta atgcattcct gcaatgctaa catgctttca taaatatact	2820
tatgatttta taggtgtttt cctcttagcc tcaagaaggg agatgtttta gtctagggac	2880
tgtcataaca aaataatata aactgtgtgg ctcaaacaac agaaatttat ttctctcgag	2940
tcctgtaggc tgaaagtta acattagggg gccagcatgg tgtggctctc ttcttggtt	3000
gcagatgtca ccgtctcact gcatcctcat gtggcctttg tgcaatgcct ctgcacaaag	3060
agagagaaac agcaagttat gtgccgtctc ttctttagg ctactaatc ctatggaatc	3120
agggccccac ccttgtaac cttaattatt tccacagagg tcatatgttc aaatacagtc	3180
acattgggag ttagagattc aacatataaa ttttagggga atacaaacct tcagtctgta	3240
acagggaaca gaggtgctag aagaaatgta gtagtagcaa aagggtgggt ggggagataa	3300
taggagcttc atatgtcttg ctaagaagtt caggctttgc ttttagcaca gccctcttc	3360
tgctagcttt gttcaggcca ccatcacctt tgatgcagat tactacaaaa gccaccaag	3420
gtgtatgaaa tgagaaaaca catacttaaa ttaacccatg gccaggtttt gctggttaact	3480
tgggaacaca gtaagggaag gaaaataact ttgttgtgga gagggcaaac caaggaaggt	3540
cagaaagaca tagagacagt ggaagctcaa ggaaccggag gttgtcccaa agtcataagg	3600
cggatgtaac gagagcacca gaacaacaga gctggaagga aggaaggaag gaaggatatg	3660
tttagaaagc aagatgttcc cagatgaata ctgcagaggt agtataactc agtgcaacga	3720
ttgtactggg taggcaaggt aggggtggcag ggaaagtcac tagcaaacat ggccacaata	3780

atggaaacta aactttgtgt ttaccatgca ccaggctctg tgctacatgc tttctgtaga 3840
tcatctcatt taatcctcac acagggattt aagcctgaag tttttttcag agaagacata 3900
caaatcccaa gtaagagcat gaaaagatgt tcattagcat tagtcattag aggaaatgca 3960
aatccaaaga tcagggagat acgacttcac accttcaaga atggctattc ccctataaaa 4020
taatatacta aatcactctt tgaacataa cattttccac aaactcacct ctttgaattc 4080
acattttgct ttgccagtgt tttatgttat tcagacatga gtttgaatct cagcatttct 4140
acttctgagc cctatgacct atgaaaagtg acaatttctc tgaatatcaa tttcccaatc 4200
caaaaaatgg gaatacctct tggagttgtg gtgaggatta aataagataa tgaatataaa 4260
gtgcctagta tagtaagagg catattattg ttcaataaat gctagctgtg aatgctactg 4320
ctatagtat ttcattcact gtagaaatga ttaattttaa ctccctcaag taacttacta 4380
aagagtgaag gttctccaaa aggcttgaat ttactgcaa acctctttcc ttgttttgct 4440
cttttattca ggacaatata ctttgctcag gtgtgtcttt gactctcaag cattgcttta 4500
aagaatgacc agactttgtc atataccatt cttaaaggag ttcaggacat gccaccccaa 4560
aatatgctgc tttcatatat tgattatttt gagctgatag cacttcaaca gcaaagcaa 4620
gagagacttt ttcggaactc agctgcctaa aaacagatcc tccaaaagaa attcaattgc 4680
catatatcct ctccctggga gttttattaa tcagcaaaga tacactttta tcacaaaaca 4740
ggaaactgga agtcaacacc acctacagaa aaactttttc ataaactgtt atatctccca 4800
ttgattcttc taaaaaccca ttcacttttc ccagaaatca tttactccac cctaagtgc 4860
ttacatcacc cctccccatc cctgttaac acagtatgta agctctcaaa tctcttttag 4920
gagtattcat tattttcctg tggcaccoca tgaacataat attaaaatga atatgccttc 4980
tctcttatta atctgcctgt tgtcagttta tttcatagac acagctatcg aacctaggag 5040
agtacagga aattctttcc tccctatat gcttaaatca taagattcaa ctgacatcag 5100
caatgaaaca aacattcaga aaactaaatt cagagacatt atatctggt tttctgaggg 5160
agaataataa gaactaatag gcattgtcag gtaaggcaat acagatgttt ttaattaaat 5220
gatgtttttt caccatgacc cacatcttga aaacaccttc ctaattaaag accagtagtc 5280
ttgagtgatt ctttgaagaa catttttact tccatttggc cttttcttta ttgagttgat 5340
gaatctgctg gaattagatt cagcgcacat aacataaaat ccaaataatg ctggcttaag 5400
ttcataagga tttattttta tcttatgtaa agaagtccgg acgtaggtaa tccagggctc 5460
gtgctgcatt ttcttaccat tgggaaccca ggctcctctg tattttctact gtattaaagg 5520
actgtcacc tcaagttcag ctcatggtac agatggcggt tggttcaaga tgactgctag 5580

aacatctgtc atcatgccc aattctagcc agcaggaagg aagaaggag gaaggacaga 5640
ggggctcata tcagctcttt gtctcctct taattggcat tgccagaatt cccaccatt 5700
agtttttcct tgcattctca tggtaacgtc taagtgaata aaagcctggg aaatattttc 5760
atttattatt atttttttaa cccaggccca ttggccagaa taaaaccagc ttctttttcc 5820
taaggaagaa ggagaaatag aattaaggaa cagtctctgg catagcttgt tctccatact 5880
tttacttcca tccagaactg agaagtttag acaacaaatc tttttttgtc caccatttgc 5940
ctgcttttgg aatgcttctt ccagagtgtc acgctctctt ctaatgactt attatcttgg 6000
cagactgcac ttttcacacc ctccataggc tgcagtgtc cattggatga ttgtccatt 6060
atgtagataa gttcctattg ctggctgttt atgcttttct cccattttta ggctatttagc 6120
aatcctgcaa taaacattct cataattaaa ttgtatgag aattctgtat tatttattta 6180
gaaaaagttt atagaagtgg acttactgca tcaaagaaag gcacttttgt ggcttttgaa 6240
ttggctaatt tcttttttgg ttaccaagt taaaagatga tggaaaatgt cgggggacaa 6300
atgctctaga acccgctagc atagacctac actaagtctt attcaaaaag gtccctgtaa 6360
acaaaaccct agttgacttg cagtagctac agaatacata ctgggttttt gagagcagag 6420
attctctccg ttctccctt agtgatactt agccagctac tggtaatttg cttattatca 6480
aaaggagcgg ccaggcgagg tggctcatgc ctgtaatccc accactttgg gaggccaagg 6540
cgggcggatc acgaggtcag gagatcgaga ccatcctggc taacacggtg aaacccctc 6600
tctactaaaa atacaaaaaa tcagccaggc gtgggtggcg gcgcgatggc aggcgcctgt 6660
agtcccagct actccagagg ctgaggcggg agaatggcgt gaacccggga agcggagctc 6720
gcagtgagcc gagatcgcg cactgccctc cagcctgggt gacagagcaa gactccgtct 6780
caaaataaat aaataaata ataaataaat aaataaagca tttttttttt cacttttagc 6840
agcgcatcag agaagaaccc acacttgatt gtcttatttt tccccctaca atgcccagag 6900
cagtattcac acattcatga aaagaaaaca catcacttag aacattcgaa tagagcacgt 6960
gtttttattt tgctcttcta tgttttaact ccacttctcc acaattttga tcccttcaa 7020
aaaaaatctt aactaattat tctctatata ttctattagg aatcttgggt gtagacttaa 7080
tcttgagggt agaaagaaaa tagatcttga agccactatt ttggcacatt ctgtaattct 7140
tagaatttcc attctaaca tttctggat gaatttaata accgtttata ttttatgagc 7200
taaagtgcc attaaaatta ttcatctacc atctttaaga cttacttgtc tgttttatcc 7260
aatgcaata aaaatgcgga actaataaga gtaactcccg atgagtgttc aacaaagaaa 7320
agaaagaaaa catatttttt ggccgacatt caaatcattt cctttttctg gtagggccaa 7380
atcttctcat attgggaaag gaaattagct ttgcttcaaa gcactttctg aagagcaatt 7440

tactaatgag cttaggaatt ctctgccctc ataaccctcc tatacattac ctgaacgcag 7500
agaaacttga gacgttcagc aaggaggagta aggccaggag tgttgaggcg tccagggtccg 7560
tctgtggagt tcaactgccac cttcctgtgt ggatttcagt ccttgtttct ctggattttg 7620
agttgcattc tcaatacatt atttttatct ttattcatag tgaaaatatt gaagactgaa 7680
ttttttctga attcatacaa tttatatgac atgttatttt ttattttcta aatattttac 7740
tgttgaagta gttcctcttt cacccaaaaa aactaatagt ttttgctgtc tttttttaat 7800
ttctacgtga ccatttcttt aatattttgt aattcacatt taattttatt gccttgtagt 7860
caaggaatat ttctgaaagg acttctgctc tttggaactt attaagagtg tctttgagg 7920
ttaatatgta attattatct taaaaatatt taatgtatac taaaatgtaa gacacaattt 7980
attgctgtgg aattaaacaa agaccatcca aatgagaaca agcaatggct atttactcag 8040
agtttgggag tcagcccaca gtttcttgtg ttgacagaga ctcaaaggc tgagtggaaa 8100
agctttgcaa gggggaaaaa aaggcagggt ctcagctatg ccctgattgg aggctgctgg 8160
cctaggggag ttgacctgaa gcacagcatc ttatcggatt gcttaagggt gcatatttac 8220
ctttctccta ttggctccta gttagaagca agggaaacaa ttaaaggagc tatcagttat 8280
gaaccaagtt ctgggtttgt tgggctaatt gctatgggtg tattgtttgg cttcctgagc 8340
tgggtgctgc aggttgtgga taagagttct attttatgta tatggctctg ccattgtcta 8400
tatgttcagt ctctcaatgc attatttgta cttattcttt tgaatttgcc tttactaatt 8460
atggctctcta tattgtctct ttttatatac atattttttg cctgtttaac gtatgataga 8520
tgactaaaag tgctattaaa aactttcaca cttttggttg gacgtggtgg ctcagcttg 8580
taatcccagc actttgggag gctgaggcgg gcagatcact ggaggtcagg agttcaagat 8640
cagcctggtc aatatggtga aacccgtct ctattaaaaa tataaaaatc agttgggcat 8700
gatgggtcac gccgtaata ccagctactc aggaggtga ggtgggagga tcgcttgaat 8760
ccaggagggt gaggttgcag tgagcagggg tcacgccact gcactcctgc actccagcct 8820
gggcagcaga atgagactcc atatcaaaaa aaaaaaaaaa aaagagagaa aagaaaagcc 8880
tttcacactt tctctcatct tttagtattt ttgatcaatg tttgttttat acatttcaat 8940
ggtttattat ttggttcta aaggctcctg attagcatat ctttaagtga tgtttgtgta 9000
tgtatgttga tctctctctc tctctctctc tctctcatac acacacataa gtcttttcat 9060
ttattagctt tgctttaaac tttgctaatt tgatattaat attatgttta tctatcttct 9120
ttttattgca cttgtttgcc aagcacgaa ctaagcattt taaataatat cacatttttg 9180
taactgtttt ggcatttatc ttcttcattg tgactttact tgaaataatt atcaatgtgt 9240

atctctatat ttgagtgttt caatacttat caagagtttt cttccccatt cttgggttcct 9300
gtgttttcgac acagctctca actttctgta gtactttggg tagtattttt tttttcaaga 9360
agtcattgga taagattata aacacctggc atttctttat gagccttttt cttcactcca 9420
caacaataat gatttagatg ggtatagaaa tcttgaattc ataactcaaa tctccataga 9480
tgttgctcca ttttaattctg gtattttgtg ttacaaaaaa aaaatctaaa gtcagtccaa 9540
gatttctttc ttgatagggt atccaatttt gtcttttgtt tctgattttt ttttttttag 9600
atctcatgct ttcaccaggt tatgtctatt ttttggctgt ctgttatgtg agttctctga 9660
gctgccatat gctcttttga tctgctgatt aatgtgaatt cgtagggtat tttctttatg 9720
tctttgaata cagattactt accatttatt ttaatcttat cttcaggaac tctaactgtc 9780
tgtctatatg ttagtaattt tggcctttct gtcctcattt aaatatcttg gtcctattct 9840
tctacgtggg aagaataatt tatactttat taatcaataa tttgattttg attcaatgat 9900
tagcagtgtc aattctactt tttattgttt ctgaagtatc ttaaagggtg ctttgtctat 9960
gaaaagtttt aaaacattct tcctttgtat tatctgggtc ctttttattt ctgtcagctt 10020
cttgggtgtt ccttctcac aacctatcta atgtcttatt atctcgtctt tcacctttaa 10080
ttccatactg tcaaagtgtc cccaaaatgt aaaaagaatg caaaatatgc gctaaaattt 10140
cagggtttct tgagctaaag tttttttgaa tatatattct ttttctgata tttttgtgct 10200
ctttttaact caatgctagg aaatctctct ctgtctcata tgtttcaccc atctactgaa 10260
agttttttta ttggctaaca tctgaataaa tgcacatata ttttatttgg attattccta 10320
cactagctgc acaatgccat cattgtttcc tccaccaaga agagctgggt gggccatgtt 10380
cattgagaaa atcagtttct ctcaaacagg agaaaaaatg gcattgtcct atcgctatgt 10440
gaataattca actaaaaaga ataactgttc tcccagtaaa gctttttctt tcttagaaaa 10500
attgggtttc tttatattcc atgcctttgc ttcagagcac aatatgtgac ctgggaaagg 10560
catttacctt tgtaaacct tcattttctc atttctaaaa ttacaagggt tatctccatg 10620
atcttttaga agccagaact gatctgggtc taatggctct ttgattctaa ttagaaattt 10680
ccaactgcaa aagttgaggg ggattcttac atcttcagaa gccaaactaa taaacaactt 10740
ttgctgtaat gagtgtcagt cacctttatg aatcagcttg tcaacttaata atagaaatag 10800
ctggcttctg gcccatgcag aaggaatgta gaaaatagac tatagtaaca agctgactcc 10860
cattctcctg ttgttcctct cctcactcag tttccacagt caacttagga caattccata 10920
ctgtgattat ttactcctgc tttgctttcg taattcaaca aacagtatta accacttact 10980
atggcgaggg atggttccag gtgctaagga cccacagtgg atttgattac aaattacagg 11040
taaaataata aagcaataaa aaggataatt ttgctttcat catcgaaagc tcagtgatag 11100

agtaagcttt agactttatt tgccctcagca atgcagcagt ggtagagggc agaggagaag 11160
tctcatccta agccttgctc ccaaggctgg ttataggatg ctgtattagt ttgctgggac 11220
tgccgtaaca aataccaaag attgggcagc ttaaacaaca gaaatttact ttctcacagt 11280
tctagaagct agaagtcctc tatcaaagtg ttggctaatt ttattttgag gcctttctca 11340
ttggattgta gatggctgtc ttctccgttt ttttcacatg gctttttttc tgtccatgtc 11400
tgtatcctaa tcttctcttc ttataggagc accaccagtt atactggatt agggcccacc 11460
cagaattacc tcatcttttt ttcttcatta tctctttaa gtccctatct ctaagcatgt 11520
cacattgtga aggactggag gttaggactt taacatatga atttagggga gatgcaattc 11580
agcccataac aaatgcctgc tagtaactca tgggaactct cttccacatg caaagagata 11640
aaatgtgatt tctgcaagtt ttattagaaa tataggaagg aacccttccc agcagggcctt 11700
gacaagcctg tgcattgtgc tcaactgacat aactaagtta tagggtcata tttgaactaa 11760
tccatggaac aagacatgca attacactga tgagcttagg ctgaagtcaa atgtctcact 11820
cctgaaaatt cagacttccc caaaacacat gggctgaaca ggtaagatgg atacttcaac 11880
tgaaagccag ggtacttaac aaaagcagga gaaatagatg ctggggaggt aatcccaacg 11940
tccagtgcaa atgctatgag ggaaaacgac agagtcccaa acatgctgga gcttatagtt 12000
agaggaagga aaaattagat ttttctcagg gctttgggat tggggcacct ggcagtcagc 12060
gcctagattc caaagcaggt cagaagatgc ctctgcccac gtattggaaa ccttttgaat 12120
atctttgagt attaaaaaca ttttaacact gcctgagaaa agtagtagtt atgaagttat 12180
acatttagcc tatcccatc aactcagggg attgtacaca gatcaaataa attgcaattg 12240
tgtttccctc cctgatatto acataggaag gtgagtgggtg agcagtaaac ttcaggacct 12300
ttaaagtatt cattgctttc acccttgac cagctttttg aggtaaatat tcccattttc 12360
taggagagga taatgaggct cagagagggg aagtggcctc cccagggctc tctagctgaa 12420
ttgtaaccag cttgatcttc cttttgaatt ccaccatctt gcctgtgagg gaaaaaggaa 12480
taaggagcca ggagttgcag ggtcattctt ctctcaaga ggcttgagag taactaagta 12540
tctgttcaaa acacttttga gaatttggcg gcctcattac gtttctcagg agtccctggg 12600
ggagggttaca cttttaggat ccccaggaat agagtgacaa aatacagcta ccacaacttg 12660
gtcctgtgca actgtgccct tgctcagcaa ttactgactc cattctgact gaactctgcc 12720
aaaaactctt aacattaata attattacct gttctcaatt ttaataatta agagttttac 12780
aacttocagt gactcttaat tcaggtttct ttggcttact gttcttgagc aataacagta 12840
tggtcactgt tgcacagaca gaagctagac atagctctgta tagctaactc cctcatgtcc 12900

ttcaaattctt tcttcaaattg tcagtttctt ttggggcctt ctgtgaccaa cctgatgaat 12960
agagcccacc agcccagac cacagcgttc ctgacttctt ttactctgct tcacttttcc 13020
ctttttcaca gcaattgtca cattctattc tactataact tacttattta ttgaatttac 13080
cactgattgt ttgcctaacc tcactataag acagggatct ttgttttatt ctccaagaac 13140
ttagaacagt agttattcaa ttaattttg ttgaagcaat gaatgaacca aatgaatgaa 13200
tgaaagatga atataaaaat atttatatgt cctaactctc tactaagcat tatgacagaa 13260
tatccaccaa aaaaaaaga tgagaaggag aaaccattat tcttctgtag cagggttaac 13320
agaacatgag cagatggcaa acgtgagatt atcagcacct tctgggctag ttcttggaac 13380
tcagctgatt tcctgctgct tccaaaaatt cttaaagctt tcagactcct gagcaagctc 13440
ctgcaaaaca cctatttttg tgctgcaagc acctgcagta gctctctagt cattcaggta 13500
cttggtgacc ttaggaactc aagccatagc ttccacgatt ttcaatccct gatccctctt 13560
cagtctcaaa ttgctgtatc caaaaccctt cctacttgca atatattaga aaggccacct 13620
tctgaccaa tctgaaaagt tgggtagggt tgtagaggat gagtttttg attattccca 13680
attcttttga tcctcacccc aacctcctaa gactaaagct acagtggcct ccctttagggt 13740
aagaaagcag aggaaaggag cagaggaatt attttctgtg aattaactgt ttctagattc 13800
taatattagt gactaaactg ccatttgtac ttctatatt cctgttttag caagagctaa 13860
aagagtttca aagaaggtaa attgtattct aagaagagaa ttcagaaagt cctttccaga 13920
atagatgaat cttaaatgat gggcagggtt ccacagaga taggtgggga agccattcta 13980
acctagaaaa caggatgcct agaagtatgc agtgttttca gaaaatggca gacacaggct 14040
tccttggtta gagaacaaag acttcataca ggagtaggtt cataaataca agcacaagag 14100
acctttatag cattgggggtc tagtgattct caattattct gcctataagc atcaactgca 14160
agaatagtta aaattctatt gaagagaaga aatgttctc gctgtgccag aaacccttaa 14220
gaaaaagcaa aggaatttca cagagctgaa gatcaagcac ctgagaaaca agtttgccctc 14280
aagagatgct tccaaaggca agttggaagc ttattgatga aaaagtgaac cactatcaca 14340
aggagtacag gcagatggga cagaactgaa attcaaatgg ctaggatggc aagaaaagct 14400
gaccacttct atgtacctgc agaacgcaa ttggcatttg tcatcaggat cagagatctc 14460
aatgatgtga gcccagggt ccaaagggtg ttgtagcttc ttgcctttg tcagatcttc 14520
aatggaagct ttgttaagct caacaagctt agcatgctga ggattgtaga accatatatt 14580
gcacgagggt acccaaacct gaagtcatta aatgaactaa tctacaagca ttgttatggc 14640
aagatcaata agaagagaac tgccatgata gataccaatt tgattgcttg atctggcttc 14700
atctgcgtgg aggatctgat tcatgaaatt tgtactgctg gaaacacttc aaagaagcaa 14760

ataacttcct gtaccccttc aaattatcct ctctatgaga tggaatgaag agaaaaacca 14820
ctcattttat agagggtggag atgttggcaa caggggaagac cagattaatg ggctcattag 14880
aaggataaaa taagggtgtct gccatgatta tttttgtaat cttgtcagtt aataaacagt 14940
gactgctttc gatttgaaaa aaaaatacat agttaaaata cacctacaca ttcacaccaa 15000
agattctgat ttaattggta tgaagccaaa ccaagacatt agtaactttc cagagcttcc 15060
caggtgattt taatttgcag ccagtgtga gaaacactga gacgatgcac ataaaatgcc 15120
tagcccagag cctgctcaag aaatgttagc tgttgcctc tactgtcact gtttgggtgg 15180
ttgagacttg tgcccaggct tcctagggga gtgttcctcc cattgccccca cagttccctg 15240
taagtgggtgg aaaaggagac cctatagcag cacttctcaa caccggctac acatttgaat 15300
caatagagga gccttgccaa ggtttgcaa agatggctgg gtctgtctct aaccaattag 15360
aatctctgaa ggtgggaccc aggtatcccc attatttact ttccccatta actattccac 15420
aaatatcctc atatgatctg atctctattt ccccataatg attgactggt atttctcttt 15480
gccatttcag tattcacaat ctgctttgta ttaaagctgc ttatttacgt tatctctccc 15540
actgtgaacc actgcagatc agggactgga tttgttttgt tttgttttag tttttatata 15600
attgctttcc ccttgcactc agcatacagc aaagctctaa cacatgatat atacttacat 15660
aatattaatt taattaattg tcagaaacca tagatactca actataagac tttcctgcct 15720
tctccactct acttttccat ttttcatggt ttctttgttg ttgtgctttt ctatgccttc 15780
ttcacactgt agccaaagt attttattca aaattcaaat taatttaatt atggaagctg 15840
ggtaatgagc ctaataatat aaacacagt gtgcagtgag catgcagtta gttacaatgt 15900
tccatctatt tttctgttgt gtgaaggcat ccacaattta taaagtacct ggtgattcaa 15960
aaattaagaa ccactattct aagagtaggt ggaggccaga ctatggaatc ttccaacaac 16020
caggagcagg agttggaatg gaataaatta gagcactgga tcatccctga aatccctgag 16080
cagaggaaca accgaaacag agatgtattt aagcttataa gtctaaataa aagttggggg 16140
tttaaaaagg ctacttaatt acttctgaga aaaaatattt tattcatttc tcaactgggt 16200
tctataatca cctcagctct gtctgcagtt tcctttcttc tgtaattatt tttctctcca 16260
ttttcttgcc ctgatagcca cgtgctcatt tgggtatcca tcaaaagact acatttcagt 16320
atattaattc tctcactgga tggtagcgt tgaatttcct tctgctagtt tcattaaaaat 16380
ctaaatgggg ggaaaagtcc tctacattat ctaggagatc catgtgagct gccagctcc 16440
attagctggt gtctgttaga ggtggattac aagcaggtgt aatcttaagg acaccgccct 16500
tctgtgattg ttaagaagga acacagactg agcagagcgg ctactgttct gatcaggagc 16560

caccaccact ggttcctaca gtatacagaa tttcaattgt cagatctgtc ctggaggaac 16620
aatagaatac aactgtgaat taaataattht agaaaccaag gacataatct cacttttcta 16680
aatgagtttg cctcttttag gagtcagtat attcgtagtt ttgctagcac agtcctgtga 16740
tgtaggagga aaatgccccaa aaatattcta gaaaaaacat tgagaattca agttaaaatg 16800
tcagatggag cagcctctag aattcctcct ttgtaggcca tcagtaatct cactgaaaat 16860
tagacatgta ttattttataa ataataatat aatattatgc atatttttaa tattaatgat 16920
ttaggtcaaa cgcagaaggt gaaatttgta ggtgtcagaa atgaaggcca aatcttttagt 16980
gggaaggctg tgggagctga agctctgcag agcacagtgg ggatccagat gcaatattta 17040
gaggctggtg ttttcaactct cctgtgcaga agtcaatgct ccaactctag agcatggaac 17100
taggcttccc aaataagcct attgctggca aagctctgca tagtccgtga ctgggagatg 17160
ggaaatctgg aaaactaacc atctgacctt gggccatccc aggcctgtcc cggatccctc 17220
cagaaatctc aagaaacagg atctcaaagc cacagggcac tatgtgcgag gcgccacgtc 17280
cactggacac attcagggaa catcattcac atgaacatgt gtccttggtc tggttattga 17340
ttgctgtgta acaaaccatc ccaaataatta gtggcttaaa acatcagttt attgcctttc 17400
atattttctga ggggtcaagaa ttttaagtga atacagtga gtggctcatc ttttctccat 17460
gctgtctgag actcagccag aaatggcttg actggcactg ttggcactgg tgccagctgt 17520
ccgctggaag tctggcaaag atcgccagga aagttcctca gttcctcctc acttgggcct 17580
ctccagaact ccctttctgg tgctacctgg gttccaagag gaccaggcga aagctgcaag 17640
gactcctatg acccagggtc tgaaattatg aaatgtccat tctgccgat tccattggtt 17700
gaggatgagt cccagctcca gccagggttc agggagacgg gaccacacaa ggggtgcgaat 17760
aatatgaaac ttggctcatt gggggccatc tttggagact agtgaccaca cgtctggaat 17820
tgtgcaatgg gccagtactt gggtgtggtt atagaaatcc cagtgtaaag cctaaataga 17880
aacatgagtt aagaataagt caatccaatg gtgcacttca gaagtgaccg taaaactctc 17940
ctgtagaagg actttcacat attagagcat ctacaaagct ccctcagttt ccaaagatga 18000
tagatctcct ggaagataaa cttgaggaga aggtaggtgc ccactatcag agctggactg 18060
tgagctccaa gctgggtcct tttacctagt gggggctctac aggttactca tcctagaccc 18120
agtcaattgc ccctgacctg cctcaggagt gaagcagagc cagggtcagg gctaggtggg 18180
taggtgggct gccccagact taagctgtgc attgagaagt gccatagata ccaggaatca 18240
aatgaatgga gatttgctat tatttggaag gatgtaagac cacaagcata tctataagaa 18300
attgtaccag gatcacaat ctcccatctc agcccagcag cagtgatctc tggaaagaaa 18360
caccacccca catcccagc tggatatactt cacagtaatg atgtgtgccc agattaacac 18420

aaatgacaaa ttccatcata ctccctctat ctgaaggagc acacctgatg tgagaaagag 18480
catgagagga gcctggctgt cagagcagga attctgcact gtggcagggt ctgctctgtc 18540
tggctgatgc ttctgaatac agcagagata tgacggcctc ataatttagg ttgcttgaaa 18600
ggcggagcaa gaaaagttac ctagtttaga aaaatcagga aactcatata tggaaaggct 18660
atgctatccc aagccagcat tgctctggaa ctaagcagat gagaatgtga gccagaaaag 18720
tttttctccc aaagaaatgc tctttttgac agccaatcag gccaagtgg ttataagggt 18780
agccaagcg gcaaaacatt tgagtcagaa aacctcacc ttggcacctc taccaaggaa 18840
ttgtcagtca gccctacaga accttctgtc tggccttacc ttaaacttgt ggcagaagta 18900
atgccacatg ttcacaatat cccatggatg gggctcatgt atgtgttcag gccaatgaaa 18960
tgtgagttga agggactatg tcaactccag actgagactg tgaaaagccc ctattagatg 19020
gttgtccctc tgctttgagc tgtgagtttg gaggccacgt gttgagatgg tagagtcctc 19080
agatcaaagc agccctgaat acaaaatagc tgtatttga gcttgtgtgc tactgcagat 19140
accataagcc tcgcctatct tgttactcca aagatgtgct ggggtcatct ttgtagctag 19200
cccaggggca ccatgatgct caggctacca tatgggaggg cgatccatga ggagcagcac 19260
ctgggtggtg aatagcacct gggaagctga acctctgtgt tcagcttga ggtagttagc 19320
agatggagtt ctaggatgta ctacatgag aagaaagtcg agggaagata aaagtcagaa 19380
tgagggagat ggtattgggt aaattcaaag aaggaaggat aaaaaccttg gggaaagagt 19440
ggaaagaggg tggcctccag agctgagtc accaggatgt acaacaaatt atgcttgaga 19500
tgagctgtcc aggactcagt atgcaaatgg gtgatctatg cctcagtgcc cacttattgg 19560
aaaacacacc tgagatgaaa agcatgctct aaactctgtc actggagagt cttgccgctt 19620
ggaccgtgag atgactataa caaaacactg ccattcatct tccaaaagag gaaatggaca 19680
gacatataat tactagccca taaaatgatg gcaaataaca cagccctcca gacactatac 19740
ttggtgccat ggaaaaatct gaacattatt gggcaatcac tgccttctga cttcaacttt 19800
gacttaagaa gagcagaaga gtgtacatat tgggagcagg atgaagacac agagtttctg 19860
gttcaaatec tcgttctcct gctttctatt gggagatctt gcaagttaat tgcttgactt 19920
ctttctgctt caattatttc atctgtcaaa tagagataaa agcaattcct ttttcatgag 19980
gttagtatga aaattacagt ggatcatgaa gaaaagttct taggaccatg tctgatatgt 20040
gagtcctcaa tgaatgatga attcattgaa tattatgata tcaccagggt tttttaaac 20100
gctggccagg tatccctca caatccttta tctctcctcc tctgtaaatg tccaacacag 20160
gccacttctc ccctttgtac acctgctgtg caattcctgc cactttggaa tgggagcatc 20220

acagcttaag ggtcctaatt cctttctcag cctctgggta tcccaagtct ctaacttgcc 20280
attgcctgct ttgcaccaag accagagtgt caggggcaga ttacaggagt cctacagtca 20340
ctgctctgct aatgagtgt tcccttagag gtgaagaagc gaggcagtgc ctgctttgga 20400
atgctttaat aaacaccaga aagtcctcac tgccagggat attttactca tttactttct 20460
aaagcctaaa attctagaag ataggacaca ttcatttgct aaaatacaaa atttccaggg 20520
agaggatgta taatgagtct tttttgtttg attgtttttt gagacagagt ttgctctttg 20580
ttgcccaggc tgcagtgcaa tagcgcaac tcagctcacc gcaacctcca cttcccgggt 20640
tcaagcgatt ctctgcctc agcctcccgga gtagcttcga ttacaggcat gtgccaccac 20700
ccccaagtag ttttgtattt ttaatagaga cagggtttct ccatgttggg catgctggtc 20760
tcaaactccc gatctcaggt gatctgccgc ctgagcctcc caaagtgtg ggattacagg 20820
cgtgagccac tgcgcccggc cacataacaa gtcttatcag tcagtatacc tcagtataaa 20880
ttaatttatc aggtccttt tctaaagggt tagatgaggg ggaaatgtaa gatggtggaa 20940
gacttgggca tttgagaagc aattgagagg gaaaatgatc caaatgagaa ccctgggacc 21000
cccagagaca gtcactaagg agttaccagt tggcatagaa cttctacttc tgctcccaac 21060
tgtccaagct caaaatgaag ccagtaggg tcttcagggt ctcatcttgg aaaaggacag 21120
cttccaagcc taggatttca gaatgtataa tggtcatttg agtgccgctc atcagtgaac 21180
aatagcta atcagtgtgcaa attcagcagg cctgagctcc atcaccaaag gcattttcat 21240
gaggacacat cctgcttgga gaattttgag tgccatcaga caagtcaaat cgctgccctt 21300
ctctttactc aaaggatctg gaatccctg aaccaggaaa aatgccactg tgacccttct 21360
gggtgcataa gcacaggaat aaacacaata gatgggcaaa tccaactgta ccttatcctg 21420
atgacatatt ttagggcagc agtttttaac tttctagttt aactcttaa aaataatatt 21480
gagaacccaa aagaattttt gaggatgcag attaagttat ctctacagat atttatcatt 21540
ttaaaagtta aagtggaaaa ttcttttaaa atattattta taaattcatt gtaaataaca 21600
ataataaaat gttaacataa atgtaatacg ttatgctaaa taactcatat tttccaaaat 21660
gacaaaaatg cagcttgcat tgttcacatt tttgcaaac tcttcgatgc ttgtttccat 21720
tcagctgtga tatcactcat gatgccacct ctagaaaact ccactctaca cacaagagag 21780
aatgagaatg aaaaagacaa aggacagctc aggattgctg taaaaatagt tttgactttg 21840
cagatcccc gaaggggtcc cagagaccca caggaggccc tgaccctact caaagacctc 21900
tttttttagag gattgtttta ggtgattggg ctgttctgat atgatccatt cctaggcttc 21960
taaacacctc aggcaagtct ccagggccca ataggaaggt caatacagcc tcttccttgt 22020
taagggtaga cctacattga agatcaagga atcttggtgt ttgtagaata attatggact 22080

agcaaaggac aaagcagttg caggtttctt tgctgagcct ctgtgtttgt gagtctggaa 22140
cccaaagggg cagcagctga ggggcagggg ctgagaagta gcaggatctg tctcttaata 22200
gctgcaccaa aagggaaagt catgacattt ttctgagact tcatttgtaa aatggaaata 22260
aagtccaagc acttcatgga gagtacctta ttttaattcaa acatataaca tagaaagatt 22320
gggagataat catctttctc tgttttacag atgagaaacc acatactcag caaagtggcc 22380
agtacacagg aagtgcaga gtcctaattt gaagctgtgt ttgtttgatc ttaaagccca 22440
ggcttttctc tgtcctgctg gcttggtgag aaggtgtcca gcgttggcag gcagctgaga 22500
gaagtggggg atggactttt ttctggaagt ttatctggag ctaatggcat gtggtgacta 22560
tgccaggggt gtgtgagcac tcaataccta ataactgcct gtaggaggat taatagaatg 22620
tggaataata gaacagcaac ctggccagtt atcaaagcaa cccaagggac tgcaaatacc 22680
ctcacagtat atgctgtgca tgcataatgt gaccgtgcac ttctcaacag agggaaggaa 22740
ggaggctttt tccttttttc atttagacac aaggatgcct ggcataattt gctgcaaagc 22800
gtttttctct ctcaggactt aggaaaatgg agtcacaggt tagcaagcta gattggtcag 22860
agctttagtg aaggagcagg tgacctagaa aagatcttcc ctaggaaaaa tgcaataaca 22920
cacagtccct agaaaatgat gacatgctgt ctctatgcac tgctaagaca tttcagaaac 22980
attatgggaa ataagaaatg gatttttcat acagctggca ccttactaaa taaatatgta 23040
tttcatacct gttttgttaa tgaggattga aactgtattc ccatattgct gccagagtat 23100
gtcccataat cacagacaga ttgtttccat tctcttatgc aaaacatcta gatggacttc 23160
cagttcgaga ctctgtctgg ctaacagccc tcagtgcac ctgtctcatt ttctcacttc 23220
tagaagatgg gaaaaatgaa aactcattat tatggaaaac caccagtgc agacagtcta 23280
atgtcagaat cagagaggaa ctgacactaa ttgtaaggca gatggagctg cattaggaaa 23340
caccaagcca aggagccctc catgcccatt ccacacttag tttcaaaaag aaaaaaaaaa 23400
aaggagccaa ggccaggcat ggtggcgcac gcctgtaatc ccagcacttt gggaggccga 23460
ggcgggtgga tcatgaggtc aggagtttga gaccagcctg accaacaatgg tgaagcccta 23520
tctctactaa aaatacaaaa attagccagg ttagtggtg cagcctgta gtcccaggta 23580
ctgaagaggg tgaggaggga gaatcgcttg aaccggggag gtggatgttg cagtgcagca 23640
agattgcaca actacactcc cgcctgggca acagagcaag attccgagaa aaaaaaaaaa 23700
agggagccag acaaaggcaa ggcagaagca tggagccctg tgtgtcagga tctggcccac 23760
cagaccaggg gttgctgagc ttagatgct actaggctgt gaggcaagca gaatgtggac 23820
tctataccag tctccacttt ctggaagaga aaacaagtcc catggggaaa tatgggcagg 23880

atgtaggggc agacataaga tatgatgcaa tgtgtagttc cccaaataaa gccctgaata 23940
tgtgagttgg gaaaagttct ctgagtaaca gtaccatgtt gagagttgtg tgtgatgtgt 24000
ggggtttggg gtgtgtgatg ggggatgggg agtgagtgtg catgtactta actgcagtat 24060
ttaagagtcc tgaagagtca ggcaattgaa ggagaagaag ccttccattg gctagagacc 24120
ccatcagaaa cttggagttg gcctggagct ttgacctgcc tccagatgca tgtcaggaaa 24180
aagtcttatt tgttccactgt acaatgaaca gaaagcaggg gcaagagtaa tttattcaca 24240
tatctgcttc taatcatcaa gtcggaaaag aaatctggat actatTTTTT agaaaaaaat 24300
agagaaacca taaataattt ggaactgtat aagtacactt tataccaaga aaaaggaaca 24360
acatctggaa ggctcaaaag aaaaggatat ctctgaaaaa aaaattatgt atatgtaa at 24420
accttagtga ctagaaaaat atatttcaag gaaatgtttt actgtcatgc aaagttgtag 24480
gaagacataa ataggaacag tggcataagc aaaaaggcag ggagggtaaa ggaataggat 24540
ggtaagaaag aacaagaatg atcaaggagc aatgcttgag gcaataatga tcaaacagg 24600
tgtaattcat gttgcataac acccaatcca ttcttgaaa gacaaatttg agatgcattt 24660
ccagaacgtc aaaggaaagg ataaggaaat gaaaacaaca aaataggaat ggaagatagc 24720
aacgagtgtt cctggaagag aaaccggaat aatggaaaa gaagcagaat caaacctaca 24780
atagaagaaa cccttcctga gcttaataaa agacggagga ggagtcttgt gtgaacaaat 24840
tcccagggtc caccatttct aagcacagta attagaacga gacccacct ccctcccacc 24900
ttcagcatca catcttgcca aagcattctt attcattctc tcaactaaca cttatttcat 24960
tccaactgca tgccaggccc tgcaattaga aaggttaaag aaaaaggccc caaagtttca 25020
ggtagcacia aattatgtat ttataatatg cacagtgact ttagatttct ttgtaaaaac 25080
aaatatatca aaacatcatg gaaaaacatc tacagtttta agacaatagt ttgtgactca 25140
ggaatttcat atgccgccga atttatgtat gaaagcaaca gtttttaaaa agatgacatt 25200
aacattaaat gcttaaaaaat agttaagggt gttccagaac caatggagag ttggaacaga 25260
attaaagtcc taagactaaa gaaataatgg tatagaaatt ctgaaattat taaaatatag 25320
ataagcataa ttggtacaat gtagctataa tgaaatgata ttgcaaata taccagga 25380
aaagaatata tataaatgta taaatcctta actcaacct tccttttatt ttattttata 25440
gagaaaagga gggggaaaag gaggggaagg agaaggaaga gaaggagaaa gaaaaagatg 25500
gaacggtttc attagtaaaa aaaaattccc aatgaattat ttaaggatgg gtataatagt 25560
ttctaaatga ggggtgaatt tgcaatatc attacaaagt gaaaaatgtt aaagtattta 25620
agaataaaac aatgatggag ttgcttatat ggacaaaaca tacctataat gaaatacaat 25680
agaataatgt ggataaaatc catccgtagg gtaataaact gtaatatgtg gtatatgtaa 25740

aaaaaccaag ttataatcta atacttagta caattctact ttggaaaaat atgtctatat 25800
gtgtataaat agaaaataag gtctagaaat ctctgcagct aaatgttaac agttgttata 25860
tctggaggaa aagattttgc atgcttatta tgttctgttt tatattgatt tgtttggggc 25920
tcatctggac tttttagttt ttctggaagt tacatgtaaa ggatttaaac aaggctatgg 25980
tatattttaga tttgtaaggg tagagacaga attgcagggg ccagtatgaa agagaaggta 26040
attgaagcaa accaagtgag aggtgatggt ggtctaaatc cagacaatgg caagaagatg 26100
gacaaaagtg ggcagacttg agctattcta agatgaggaa ctgatgggtg tggagatcca 26160
ttagatatgg gagtgtggct gcagataatg acagttccgt atgtcagcat caacatggag 26220
ggaagtgtct tgagccacta gaagatcagt ctgggaggag ttggcaaggc ctcttagaag 26280
aaatcatgcc tcctttgagc tttaaaagat aatgaagagt cagccagctg gagaggtgag 26340
gaaggccatt ttagcctgaa gtgagtgtcc cattatttga cctgtgtcac agtggccttt 26400
cccagccaaa ttgcaaacag agtcaccaa ttcgaataaa catttttaaa gggagaagaa 26460
agcactgctc cctctgaaa ctgtgaacaa attaagccag agcaaatccg tccatcctgg 26520
ctgcagcaat cccaagataa tgctgctcag ttctaggctg cttggcagaa aggggttatt 26580
ctctgttatg tgccctattt actcatgaat ggagaacttc tgggaatcca ggatcagcca 26640
ggcacagcac agttactatt tctgggaaga tttcatttga cccaattac ttacaatagc 26700
agagtgatgt ccattgtcc aataataatc accatgacct tccctgggtc taggtgtcct 26760
ttacccaaag acctgcttca aaataccatc taacagattg gagctcctgt caaaacaaga 26820
agagtcatca cactgacagg atctccctga ggacaccaag tgccataagaa tttgaacaaa 26880
gcaccccttc tgtgcaggag gaagggggca caagatttct aaatagcata aataaataaa 26940
taaataaata aataaataaa taaataaaga tgaattgtaa ttatctatgt ttagcagaca 27000
tttgaagtag aaataggtea ctgggatttc catttcaa atctattgaa attctttacc 27060
aaaagaattc ttttgcactt gatctatcca gccttttttt gagaggggga tctcactatg 27120
tttccaaggc tagacttgaa ctctgtgtc caagcaatcc cttcatctca gcctccggag 27180
cagggtcttg ggtgtgcacc cccactgcac cccgctcagc actttttatc agctataatt 27240
tataacaata atagtataa taattataat aaaataatat gaatgtttgt taccatctgt 27300
caggctctgt tatgaatata tcaactgtct tataacaatc acatgaagaa gctaactcaa 27360
tttctacaga ggtcagagag aaaaattatc catttggttg aaccgaaacc catctatctt 27420
ccacactgag ttagtctttt attttaataa aatatgttgt tatcaagaaa ggaaatcatc 27480
cattatactt tcactttcac ctttgcttcc aaaatcacc aaacctgaat tactagctca 27540

tttctccaaa agcataatta atattgactt tcttttttatt atctcctata actagatatt 27600
gaaaaacact atagctagca taaaggaaaa cttctgaaca gtgtagccat attgccacag 27660
aaacagattc agaaaaataa acctaagaac tctggagaaa gaatgcatgt tatataagtt 27720
tctccctaga gtttggcagc ctttgttcat agttggatgt aaacttattc atttctaccc 27780
ataagcccac atagctttac ttcctgacat ctgttctctc ctcccaggat cggaaagggt 27840
tcttccttct gataaagtgt gaaacatttg tgaggaaaag gtgggttcag agagggagag 27900
aaagacgaca gcctgattcc tagatctgag gtgggacgca ggaagatgga tttcttttaa 27960
ttcccccttc aaagttattc ctctcatgt gcttttggga ctatgctttg agaatctctg 28020
gagtggagta tctggtggtc tgcacatagt ggtgaagagc gtggagtcac agaatttggg 28080
gtaaaatccc agctctatgt ctaatgagct ggggtgaactt gacaggtcaa ttcactggct 28140
aagcctcacc ttacctgttt ttaaaatgga gagtggatga ataagaaaag tgcttagcat 28200
agtatctagg ttacagtagg tgaccaataa atgggagtga taactaccat cattatatta 28260
catggaaagt acaaaacaaa tagtatcata gtatgtagca aagtcctcat tttgttgga 28320
aagagtgatt aaataaaacc gcaaagtctc gtgcctgtgg ggtgagatga gggagatgtt 28380
caacggtaat agacagcaaa gtgttaccaa gctacccaaa tggatgatcc atttaatttc 28440
tacatggaaa aggagcaaaa aaggaaaggg acctctgaaa ccttctaag taccacaca 28500
cacacagact cacacacaca cacacacaca cacacagaca cacacacaca ctcacacaca 28560
cacagagctc tcagaacact gctgccagag ggtgaatgca gatacaactt tcataaccat 28620
tgctgaaatc acatatattc caagaaggga aaatacttgt aaaactgaaa cttattaatt 28680
aagatactat agccaaaatg gatgtcacia tgcttgtctg aataccata ataacattta 28740
tgtaaaattt actatgtccc atcaactgtc ccaaactctc taccgggata agatcacgct 28800
atcttgacag ccactctagt tggttaagtgt tggttttgag acatctacaa ttgaggatac 28860
tgaaggacag agacactagg taatctaccc aggctgcaca cctcatagac agttgaggca 28920
agactcatat ataagttaac gcatcccaga gcctgtgctc ccaaacacag cctcatata 28980
gaactgctcc tgtagttaag aacatagtga ggtccccaga atgtgggcaa aatctagtgg 29040
agagactctt ttccgccccca taccagctgt tttcaccaga ttgttagtaa aaagaaagt 29100
gatcaagaga atgagaagac aagccacagg ctgggagaaa atattggcaa aagacagaca 29160
tcggataaag gactattgat atggtttggc tgtgtcccca ccaaactctc acctgaact 29220
gtaataatct ccacgtgtca aggggtggtc cgggtggaga taattaaatc atgggggcag 29280
tttcccccat actgtttctg tggtagtga taggtctcac gagatctgat ggttttataa 29340
atgggagttc cctgcacaa gctctcttgc ctgccgccat gtaagatatg actttgctcc 29400

tcccttcacct tccactatga ttgagagacc tctaaatcat atgtggagct gtgagtcagt 29460
taaacctctt tcctttataa attaccagg ctcaggtagc tctttattag cagcatgaga 29520
acagactaat acaactatgg tacaaaatat aacagaattc ctaagactca acaataagat 29580
aacaatatgc ccaattaaaa atgggtaaaa gacttgaaca gacacctcac caataaagat 29640
acacagatgg caaatgaaca tatgaaaaga tgctccacat catatgtcat cagggaatgc 29700
aaattaaagc acaatgaca tagcactaca catcttttag aatggccacc atccagaaca 29760
ctgacaacac caaatgctgg tggggctgtg gaataacagg atttctcatt cactgctggg 29820
gggaatgcga aataggactg tcactttgga agacagtttg acagtttctc acaaaattaa 29880
atatactctt accatataat tcatcaatca cacttcttgg tatttatcca aatgaattgg 29940
caactcatgc ccacacaaga atctgcacat gaatgtttat agtagttata ttcataacag 30000
ccaaaacttg gaagcaacca aaatgtcctc agtagatgaa tggaaaaaca aactatagta 30060
catccagaca atggaatatt attcagcact aaaatgagct atcaaaccat gaagatatga 30120
aagaaaatta catgcatatt actgattgaa agaagtcaat ctgcaaagac tacatattat 30180
gattcccaag tatatgacat tctggaaaaa acaaaaacta tgaagacagt aaaaaaaaaa 30240
aaaaaaaaatca gcagctttca gggattaggg agggggaggg attaatcagc agagcatgga 30300
ggatttttag ggcagtgaaa ctatttcata tgttattata aggggtgata tgtgtcattg 30360
tacatttgtc taaaccata ggatgtacaa caccaagagt gaaccctaag gtagacaatg 30420
gactttgggt gatgatgatg tgtcaataca ggttcatcaa tagtaagaaa tgtaccattc 30480
tggtatggga tattgagagc agggagtggg ggttgtgcat gtgtgggagc agagagtata 30540
tagaaaactct gcacttactc gattttgcca tgaacttaaa actgctctga aaaataaggt 30600
gtttttttta atgatgcata ttgccaatt atattttaat aaagctggga gatcaaaagg 30660
agtggagggt atcttcaaac acatgtgcag ccttttaaaa ttcattttata aacaggatct 30720
ccaggagggt gatcatgttt atcttgttca ctttttaaaa agtcattttg aaaaagcaat 30780
aaattcaaag gaaattgcaa aaataaagag gcctgatgta ttattgacca agtttccctc 30840
aatggttttt gcataacatt taacgtaact ctagtgtgat atcaaaatca ggacaatgat 30900
gtatgtacaa tgtgtgtata gctggacacc actttgttac ctggataagt tcatgtagca 30960
accaccacaa tatcttggtc acttttgtca tccggtgctg atcatgattc tcaacaatac 31020
tcttaagtga ctcagtcaac aaacaaataa gcaagtgaat gaacagatga ctgaatccat 31080
gacatcagta aagagcatgt ggttcttctg atttgcatat gcactcagtc aatctcacag 31140
atcattttta tttgtataaa gccttgagct gagagcagtc tattaaatac aaggattggg 31200

agcacctcac cacacattaa agtgcctttt cttccacttc aacttgacta attgagccta 31260
tgtgagtgtgta actgccattc tcccaaagat cgttattgaa ataataaac ataaaagtgg 31320
aggaaatgat tttttaaaaa tccctgatga taggaaagag tgctatccat atgggttaaaa 31380
aaaaaacctg aggcttttctt gccagattca gaacagggtga gatggaaaac caccatctgg 31440
gaagaaaaac aaaaaactgc ctttattttt ttttttgagg tgtaagtttt ttttattatt 31500
attatacttt aagtttttagg gtacatgtgc acaatgtgca gtttagttac atatgtatac 31560
atgtgacatg ctggtgtgct gcacccatta actcgtcatt tagagatggg aagagatacc 31620
catcaaaact tcatttctca cccccaacc aagagctgga aataaaagggt gtacaagggtg 31680
ccagttacag aggaaatgca ccagaccctg tcgctggaac agggatcag aacatcctcc 31740
caatctattc tgggccaaagg aggcattgctg gaggggggct gggctgagct gggtagatga 31800
tgtgggggtgc acgctggcag agccagttcc cacctgagcc agtgatgcta actagagagg 31860
cgtgcagtgc ctcaccccc gacaaagctc agaagctcca ggtagagcag acgctcttgc 31920
cttaatcttt aaaaataact ggactttgat ctggccaaat ggaactggag aaaattctcc 31980
ccctcctttt gtctgttcac ctgaggtttt tcaggaacca tccaagcaat taaacaggag 32040
tgacattatc tctgaagaaa ataatttcag ttaccagttt ggtattatga aagacccttg 32100
ggaataaccc ctctgatggg cttcagaccc atgatatgta tacgtacttg ttttcttctc 32160
acaagaaact gatgggcatg aagattgcag tgactgaact gtctccctgc gtttcttctc 32220
acataagctg taaatgggaa caaagagaga agattaaatg tctgtaaatt ttcattgcag 32280
cccctttttg tttagaaagc atttctgctt accttttata tttctattaa ataatagaga 32340
gaaatttact cctccagctc ccaagagaag aatgaaagtt aaattagaat ttatagtgga 32400
cccacaattt ctgacctag aacattaata caggaatcca aaccctccc agctcttctt 32460
tgctatagtc atcacccact cagaggttca gctctgaaca aacccttct gagcctctcc 32520
caggaagcct tatcaaccca cagtccagaa ccaccttcac cgtttgagaa cctcacaag 32580
ggtagcttga gtgctagtga tgcttacaga ctgcttgccc ttgtacttga gatggagtga 32640
agacgagggga aacccttga ttgttctcag ttttccctt tgtgccttga aatttccatc 32700
tcattgtgaa taaaaaagtc aaagcatatt caatcccaga gaggggttga atccctgaat 32760
ttcacaagat tgaaagatca tgaaaaaatt gacccctct ggaaagatga gaaaaatcaa 32820
aaagatcaat gaacacatac aaagagtctg tatctatggc tcttattaac agtgactaaa 32880
gaatttatct gagtaattta gctaattcaa ctcaataagg aaacaacata ataaaagata 32940
taactgaaaa aatgaaatta tcaatttggg aatgatatga ttaagaataa aaaacaagat 33000
tgtttttgaa aaatcattgg gcctaataag acatgtaatt aaagaagatg attataaaac 33060

caacaaatat aaaaagcaac tagcttctct gtacactatc aataaccaat tagaaaatac 33120
aatcttaaaa aacaaaagat taaatattca agaacaatca aaatacaaaa tgcatatgaa 33180
caaacttttt tgtttgtttt ggggacagag tcttgctctg tcaccaggc tggagtgcaa 33240
tggcacgac tcaactcact gcaacctccg cctccccagt tcaagtgatt cttctgcctc 33300
agactcccaa gtaactggga ctacaggcac ccaccacca cacctggctc aatgtttatt 33360
ttttagtagc aatgggggtt caccatgttg gccaggctag tctggaactc ttgacctag 33420
gtgatctgag tgtcatggcc tcccaaagt ctgggattac agacatgagc caccgtgcct 33480
ggcatgaat aaattttaaa agaattgtga tgacttacat gaagaaaatc gcacaagttt 33540
actgagtgc attttaaaac ttttaataatt gaaaatccat tcaatgttcc tggtaaggga 33600
acagaattaa tggtcagaaa caaatcaa atataaaga acctaggatt tttttaagt 33660
catttgaaat gaatagagaa aaaagggatc atgtctccag cacattgatg ttctgtctt 33720
tctgaacaca ctttaactctg tgttgctga agacctctgt acttgctgtt ccctctggcc 33780
tgaacagtct tccctcagag cctgtttctg atctctgctc aatgttctt taaaggatag 33840
cagtgtcttg gggcttttat aactttggg cttatctatg gctagcaaat gtaggacata 33900
gggtgttaga gtatgcaaag caggaaggct ctcaatgggt aaatatctgt ttatctgggt 33960
atgttttaaaa caattggatg tgtaaaaatt tgagcgtggg gctagtgggc atttaagcta 34020
acagatctca gcctactatg aagaagtc ataccaggga gtaatatata gaagccttct 34080
ggctcatgtg catgacacac ccaccta atctcaagctc ctccctcttc attttttatt 34140
atagcaacta tcactctgtg acattataat ctatatctat ttgggtgtgt gtctgttatt 34200
tgtttttctc attagaatgt taactccatg gaggcaggaa ccatgtcatc cttgttgact 34260
gcattttcag taccaaaaaa aagtgagtgc ctggacagaa tggaaacct ataaacattc 34320
accaaggaa gtaggagagg agaggggcag gagagaatat agtaaaatag gttaaagttt 34380
ccaatatagc catgaaatta tttttaggaa aagaatttca ataactcgca tagaaaaaga 34440
ttttttaagc atgataccaa ggcagaaatc cttacaaaag actaacagat ttactgcaa 34500
aatataaaat ttacaaaatt taaattgtaa agattttttt aatacaata gaattataaa 34560
cggatttaaa tgtttagaat ataaataata ttacatagc tgatcaagag ttaatatact 34620
cactaaaaaa ggactttcta caaatcaata agtagaagat aaatacttaa tacaagaata 34680
aatacaaata taggaacaag taagatacac acaaatgca attcaccagt aaacctgttt 34740
taaattgtta attttgcaa aaactaaata aatgaaaata aaaataacaa tgagttacca 34800
aattttcaa taagaaaaga aagtaggttt taagaaattg gtagtgcatt aggccatttt 34860

tattaaataa acctaaatg gtatcaatga aatattttaa agtaccctg gggtacattt 34920
tgcttttttg tacttttctaa ttattccacc aatatgtatt acttatgtaa ttcaaagaag 34980
tgattttaca agaaaaata aatagtatgg taaatcgctg gctcctagag agtgggcca 35040
tcttgaaaa caatgcccc gtgctgagaa ctgtgtaggt aagttgcatc tttcagtagc 35100
acaaaggaga gattgggtttt tcactacttt agatgagggc caggagctga ggggttaata 35160
gtgcacagat aaatcttgaa ggacaaatga agttttccac atggccaaag gaaggagaaa 35220
ttataggcag atgaatagaa cgtgtgaaaa gagaggaaca caagacagca tgggccaatg 35280
tagcatcata ataaaaataa gcaacacaag gccagggtgca ggggctcatg cctgtaatcc 35340
cagcaccttg ggaggccaag gcgggcagat cacctaaggt caggagtcca agaccagcct 35400
ggccaacatg gcgaaactct atctctacca aaaaaaatta gccagggtcc acggcattcc 35460
agcctgggtg acagagttag acttcatctc aaaaaaaaaa aaaaaaaaaa aaaatagcaa 35520
cacttggaat tctcagtttg tagaattcac aactgatact aaaagtctat catgtaacag 35580
ggctccggtt gaagttcacc tcactgggta tgagttaaag acagggtttt tgatgagact 35640
gacacagaaa atgtcagcaa gaaaagaaag gacacatttc tctgatctcg cactctgggg 35700
aatgctggaa caggctcatga gtgtaccttc tgcaagaggg gccagctcc ctttctcact 35760
catcaagaag ttgctcaact ttctaataa tcagtagata aagcaagttt accagttagg 35820
ggaagctagt ggtgggggca ggaactggac cacttgggca ccagagggtg tgattttgat 35880
gtcatttggt tagctgctta ggaaagccac ccactgccag cttaagattc aggggaagtt 35940
gagatcaggt gcagccccac ctctgacat ggggaaaact cattgtctc actcattcat 36000
catgcaggca acctggggag aatttccacg gttgcttttg gactgccta tttgtgatac 36060
cagagctcta tgacatttca ttaattttaa ttttactgag gctgaatttg actcactcat 36120
gcttgaaagg tagggtgtat aacacatcac taaacaagca aatgaggctg tggaccacca 36180
gcagtcacat ctcaactaga cttctgttca ggtatttatc agcacatatg catgctcact 36240
cacaaatga gtagccacaa aggggtgcta attcctaag atgggtgggg ccgttaagag 36300
tgaaattggt agttcttgga cagaaaagag agttctggat aaatgacaaa gaagaaaaat 36360
attaacacat ctaaaagttt gcatgtgaat acagcacttg gtaaaagtaa aggtgatttg 36420
tgaagctggt ttaataattt cttgaaccaa ttcaaaaatt gctcatcaac ttaaaaatta 36480
gacaaatgaa tgatgaaacg gtggttaaca gagtgtaac aggtcactga ctcatttaag 36540
aatccatgca ggccagggtc agtgggtcat gcctgtaatt tcagcacttt gggaagccac 36600
agtgggtgga tcacgaggtc aggagttcaa gaccagcctg gccaacatgg tgaaacccca 36660
tctctactaa aaatacaaaa attagccggg cagcatggcg gttgcctgta atcccagcta 36720

ctgggaggct aaggcaggag aattgcttga actcgggagg tggaagttgc agtgagccac 36780
gatcacgcca ctgcactcca ccctgggtaa caaagcaaga ttttgtctta aaaaaaaaaa 36840
aaaaagaaaa agaaaaagaa aaaatgaatc cttgcagagc aaaagtacta tcaaggcttt 36900
gcaatagcct atcctgcaaa ataaaagtta gttcaaattt ttaggaaata tttttctttg 36960
cagaatttag attctctttg tgaagagacg gcttttcatg atcacacctg tatataaatg 37020
tggactcaat aggaataatt aacatcattc atcactcttg gtcagtgcta tcatagagat 37080
gtaatgcaat atttttgttg aggatctttt catgacaaag tacactcatt gctatttaaa 37140
ttacattact tttgcattta tagaatctgt gtggaagtca cagtttttcc attataagtt 37200
atggccttct atgggataaa attacatgac tttccagaaa atggagtttc acaactgtct 37260
tgagatattt ctcttacgaa tgccgaggac tactttggta ctgctataga aaactgggag 37320
ggggaaggga cagagatgat attgaggaga tggacaatga ggttgtcttt aaagctatgc 37380
agaggttagg gtcacagagg agtcatagtc atgggagtga cagaatcaga ttggtgtttt 37440
acaaatattc ccctggaaat gaatggagga cagatgggaa agaaaccaga ttggaggagg 37500
caggttgaca acttaggaaa cgaatttaat agccctgaaa agatgatgaa gatctaaagt 37560
taggttcaaa tccagtggtc ctcaacttta gctagatatt agagatggct gaagagcttt 37620
tgaaactaca gatgttgaat atcattccag ataaattgta ttttaagagtg acatctgggt 37680
atcaattttt taaaggtccc cagatagttg tgatgcaagg taagaattga gaatgctact 37740
ctaaacaaaa caaggcaata aatgtggata cattcaaaaa gcttttagaa aacagatgat 37800
gactgatgga tgggaatgga gaaacaggca ggccctcaact tccagatatg tgtcttgaa 37860
agcatggatg atagggtcga aaatattaac acatttcaag ctttgaaaaa aaacactaag 37920
gtataagttc ctcaaactca acctgccctc tccaggccca cagactggga gtggcaatca 37980
gcaaggaaaa cctctattac tccaaaaata atcttttctt tggttgggaa aagtaagtct 38040
tctcaaaact cacaagagaa aaagcactaa aggtgcttgt tctgaaggga gatctggatg 38100
ttgaaatagc agcatcccat tatgtgagat tagattgatg atttcttcaa tactgagatg 38160
gagcgagtgt cctcagagga cagtgtttct caagaagggt gtctaaatga ccacaaagtt 38220
tgataataat taacagattt cagtctccat gaaaaccogt gttacctgca agtgtgttct 38280
ctgtcacctc ctattggcct tatatttttc tctcttagct tgccctcttg gctgtacttt 38340
tttctgagat ggtgttatgc aatctcccca caaggacct cagtagcact aggttacac 38400
cacccttaga gtgagtaacc agtcactatg gccatggatg gactactctg gccagcccta 38460
ggtcataaga taaccctggg taaggagaaa gagagtcacc accacctaaa ccatgtggat 38520

gaagacctgg ttgggggaaa gataattgag atagtgttat tagaagaaga aatattttta 38580
gggtaggcaa aaacaacaga tgatgactaa aacctcaaga gagtctctcc actcccttgg 38640
ccctcatcag ggtcatgcac ctgaggaatg attcatgggt gactctatcc cctccctgggt 38700
ctccattcat agccatgtgt ggaggcagag atgaccatgc tggagaagag accaccccag 38760
aagacacagc taaaataaag gcaaatttgt tgtgccattc ttggcagggc catctctctt 38820
gtttattatc ctggagccct cttcccctcc tctgttttgc tgaccttaat aaatgaatac 38880
atcaatatat gtttggaat acatctgttt ctacatacct tatgccttcc ctccaaagaa 38940
taaattaagt gattagtaga atataagagc tgaaaatgtg agagctatct gctcagaatc 39000
cagataaaac agaatggggc ttatgaaaag tagtgcttga ttagttagtt tcaaaaatcc 39060
taaaacagtg ttttcaaag ttcatttctt tgaaacagta gttaagggtt tgatactctc 39120
ccctgaggaa gtcatctaaa actgctgata aagcttttgg caaagagctt ttcattacgg 39180
ggctatttat aacaacagaa aaccgtaaac aacttaata tccagcattg aaagaacagt 39240
taaataatag atggtaaadc cacacaaata atattatgtt ccattaaaaa ctattgacaa 39300
taaaattgta atagtataaa aaaatctcaa catgaataaa aagcagaata gaaaattata 39360
tatgcaccat aaactccagc tatataaaaa agtgaagcct cacattaaag acagggatga 39420
gaaggatcat ttaataatgt aaataagagt tgttttgatg gtgataaaac atataagatt 39480
gtcatctccc aactttctac tttcctctag tttttatatt ttgagtaaga ataatttaca 39540
caaataactt acccatgttt agttctcagt cttcatcttg gttaagctgt cagcagaatt 39600
tgacacatct gatcactccc tectccctga aacaacctct ctgccctccc cgtctcctct 39660
agtcttcata taataataca atataataaa ggtatctatt tatgtgtttg tctgacacac 39720
aatactgtga attcctttac agcaggtact atgctatact catcttcac tgttttttct 39780
ctccttacco agattgaatt ccacagtcatt ttttataagt ctcttttgca catcctcaac 39840
ttccttcacc ctgcctcttt tcatgatatt actggcaaaa ctacaaccct ggtaaaatcc 39900
tgctttttct ctctttgtac ccacactggc gaacaaagct ggaaacaaac acacacatgc 39960
acacacacac cctacctact ctcttgctt acttacttac ttactttatt tatctatcta 40020
tttattttag agacagcatc tctctctgtt gccaggctg gaatgcagtg gtgccatcat 40080
agctcactgc aacctggaac ttctgggctc tagtgatcct cccgcctcag cctcccaagc 40140
agctaggact atggatcatc agtaccacaa tcagctcatt tttgtatttt ttatttttgt 40200
agacataagg tctctctctc ctacttttaa taacctagaa ctcaaataag tcacagatcc 40260
taccagcaa ccatatgctt attcccttac ggttcattga attctccaaa aattatcatt 40320
agaatttttt aaatattcgc aattttcaca ggataagaaa gcagtatttc atcctacatc 40380

tgcgaactgcc ctgaactctt tcttactttc tctctcttct aacttttttag cacctcctct 40440
ttcatcctct cctcaaccag atgactttgt ttctactcca ctgagaaagc tgaagcaaca 40500
gaaggggaatg tctcactcc ctggcctctc tatacctttg ctgccaatat cagcattcat 40560
atccccctgcc tctcacctga aaagtatgag taaactctcc atgtgacaag gttaaacctat 40620
cttctcatgt tttagacacc accacctctt acctactcta gatcttcgct ccaaaaatct 40680
tcccccaacc ccgattaagt tttctcttct tgetgattat taaaatccac atttttcatg 40740
ctgttctttc tcaaactctg aaaacacaaa cgaaaaagaa gaagcttatt catcccactt 40800
ttccctccat ctgcactgc atttctttac caccattag aggaaaactt cttgcagtaa 40860
ttgtccatac tcaactgtat caattcctct tctccattc ccattcaaatt tgtctagcca 40920
ggcttctgcc agtaaagtgg ctctggttta agtacccttc agtaactttc atattctaaa 40980
ccgagggctg attatttgaa cggacagtag tattcaacac aactggccac tccctcttct 41040
tgaaaatact tttttacttg gtgtctagcc accaaactct cctggtctgc ttcgatccta 41100
tgaattgctt cttctcagtc tcttttattg gttccttctc ttatttctga ctttattaat 41160
atttctcata ttaatagagt ccctaccct cagatcaatg cttggatott cttctgtaact 41220
tagggacatc atcatgtctt ggggctttta aagtgtctct acacggatga cccctgaagt 41280
gatatctctt tccaggagtt ctcttttgaa ctcaagtttc cttagtccaa cacagactat 41340
gtttactgcc tatatccac caccggaatg gaaactctgt ccaaggaaga tttttatctg 41400
ttttattcac tgcacctaga aagggtgctg gcacattacg gacatttaac aagtgttttt 41460
gaacgaaagg ataaataaat aaatgggaaa ttggagcatt atggttgaag gccaggagag 41520
agcatttttg gaggcataaa gagccaaaat caaatcaaag catcttgaac aagaaattca 41580
taaaatcttg ccttttgtgt actatagtat cttcataaaa tcaattagga aataccttct 41640
ttcactatc tctagaagat tttgtaaaag accaaaatga gatgtttctt gaatgcctgg 41700
tggaattcac aggtaaaaca atctgggcct gatgctttct ttctaggaat attttacact 41760
aacagattcc acttcttttag gggtacaaaa acaatcaggt tttctatttc ttctcaaatt 41820
atttttggca agttattttt taggaatgtg tgcattttat taaagcatc aatttactga 41880
cataaatgtg tcaatagtgt tatactatct ttttctgtct ctactgtacc tatagttgaa 41940
tgtctgattt ttacatcaaa tagtttgttt ttggcttctt ttttttcttg taatttatat 42000
cacaaatact ttttcagctg gcttgattga tctttctat tatatttggg gtttttattt 42060
gttcctgct tttgtctctt tcatctcctt tctctgatt taaattgatt tatataccat 42120
tattcttttt ctaacatctt tttattttta ttgaggtaaa atttatataa catatcatc 42180

atcaaattaa ccaagtgtac aattaagcgg cagttagtag attcaaaatg tgggtcaacta 42240
taagatctat ctagctccaa aacccaaaag gaaacttaat aggaaacctc agaaccatta 42300
aacagtcact tccccacacc ccccatccag ccctagcaac catgagtctg tctatgaatt 42360
tgcccattat gaatatttca tataaatgga atcatggtag aatatgtggc tttttgtttc 42420
tggtctcttt cactttgcat aagatgttta aagttcatcc atattgcagc atgtatcaat 42480
acttcatttt tttatggctg aaaaataata ctccattatt acatattatg ttttatttat 42540
cagctcatcc actgatggac atttgggttg tttccagctt ttggctattg ggaatgatgc 42600
tgctattaac aatcagacat aaatatttgt ttgaaacctg gtttctattc ttttaactat 42660
atacctagga gtaaaattgc tgggtcatat gataattcta tgtttaactt tttaagtaaa 42720
tacaaaattg ttttctacaa tgactgcac tttttatatt ctcatcagca atgtacaaga 42780
gttctaacat ctctacatcc ttgccagcac ttgttatttt ccttttttaa attataatca 42840
tcctagtggc tgtgaaatgg tatctctgtg gttttgggtt gcattttcct aataattaat 42900
gatggtgagc attttttcat gtgcttggtt accatttggt tatcatcttc agagaaatgt 42960
cacttaagtc ctttgttcca ttttttagtt gcattgtcca cagggatttc ccttgatgta 43020
gtactctccc cacttttctt agggatgtgg ctctctgaga gctgagctct agtgattgtt 43080
atctctcttt tggatctagc caccgagcag gtctaccagg ctccaggctg gtactgccag 43140
agtctgccc gagtcctgtg atgtgaacca tctgtgggtc tctcagccat ggatactagc 43200
acctgctctg gtgcagggtg caggggggtg aaatggactc tgtgaagtgc cttatttttg 43260
gttgtttaat gcattatttt tgtgcttggt ggcctctgcc aggaggtggt gtttttaaga 43320
gagcatcagc tgtggtagta tggggaggat caggtggtgg gcagggccct agaacaccca 43380
taagtatacg ccgtttgtct tcagctagca gggtgagtag gaaagaacca ttagttgggg 43440
acagggctaa gcattgtctga gctcagactc tccttgggaa gtcttagtgc agctgctgtg 43500
ggtaattatg gtgtggttcc caggtcaatg gagttatgtt cctgggagga ttatgcctgc 43560
ctctgctgtg tcatgcacat tgtcagggaa gttggggaaa gctggcagtc acaggcctca 43620
cccagctccc aagcaaccca aaaggccggt ctactccca ccatgccccg ccaacagcac 43680
tgagtttgtt tccaggcagt gggtgagcag ggctgagaac ctgtcccagg ctacctgcct 43740
cccagctgtg aaagcaagta gagcttcagt acttccccac ctgtggagtc tgcataccag 43800
atacacaccc tcccccaagt tctgtccagg agacttctag ttggttagca attgttacga 43860
agttcagctg gaggtttcct tctccttgta gccttttccc agtgccctctg gcagcactcc 43920
ccaaggaccc ctgtgaggca aggcagaaat ggcttggttag gggaccacac gagcccacag 43980
ggctttcctg ctgcttctc taccctgta ttctacttgg ctctctaaat tgacttagct 44040

ccagggtatgg tcagaatatt ctccgtgcat ctagaccttt aggttaccta gtgacgggtgt 44100
gtgttttgggg gtggacaatc tccctttccc actttcacag tttgggcact cacaggatttt 44160
gggggtgtctc ctaggtcttg caggaacaat ctgcttcctg cagaggggtct gtgggttctc 44220
tcggctttcc tgatttatcc ctgcagtcac tctggagcaa gagttcatga agcaagactc 44280
cacatgctgc tctgtccatc taagtgcagc ctgcaatcta gtccctgcctc ctgtccaccg 44340
tgagcctctg taaatcctgg atatggaatt cttgattgac aggattatct ttctttcagc 44400
accttaataa tgttatccca ttgccttatg gcctccataa ttccctgaaga gaaacctact 44460
gtcaatctta tggaggcact cttgtacatg acaagtcatt tcttttttgt tacttttaag 44520
attgtctctt agtcttttgg ctttcagaaa ttttagtaaca tgggtgtcttg gtatggatat 44580
ctgtaagttt ctccctggag ttcacagggc tttttgcatg tgtagatttg tgtcatttgg 44640
gaaggttttg gccattatct ctttaaataa tctttctgtc cctttatctc tctgccttcc 44700
tctaagatac ccagaggtga aagctcaggg tcttctcatg ttttttttga gagtttatcc 44760
agtccctgag atgcatttga tctctttgat ttacttgtac atattagtgc ttttgaaaat 44820
tgttattccc ccatatatct tctttcccaa cctcattctt cctaggattc ttgatctatt 44880
gtttgtcctg actgttatca cctgctccaa gctgctgtgg acaattcctt tggctttaaa 44940
tatttttttag caaatgtggc tcaggaagct ctatgacctt gggaatattc taagttaggt 45000
aaaacaaagg cacactttgt actgatcctt cagggacctt gggaatattc taagttaggt 45060
aaaacaaagg cacactttgc actgatcctt cagggagctg ccagacagtc aaaacacaca 45120
acaaaaactc ttcaagaata aagtctatac tgtcccctgt gacattagca aactggaaca 45180
gtattttcat ggccaccatt actctgggga agaggggtga caagcaggca attttaaat 45240
ctctaaccac ctcttaccac aatgcagcag cttctttctt cattaagctt cccttggttg 45300
ttgttagttt ttcactgtgt ttccctagttc tgtgaaagtt gattctacca gtttttgcca 45360
gcatttggtg ctggatgtat ggaaccttag agttccctac tctgccattg ttggcaaat 45420
caattgcacc tttttttaac ttgtcaaatt agttgcttag atagtttatt ttcactttct 45480
tcttctctaa tataagcagt taggaacctt aagtttttct caaggatttt ttagtttcat 45540
catctcacia attataatgt gctattctct caatctctct cctttttttt ttttaggttt 45600
gggggtgcat gtgaatgttt gttacatagg taaactcatg tcacaagggt ttgttgata 45660
gattatttca tcaccagat attagccca gtaccaata gttatctttt ctgctcctct 45720
ccctcctccc accctccact ttcaagtaga ctacagtgtc ttgtgtctcc ttctttgtgt 45780
tcataagttc tcatcattta gctcacactt gtaagtggga acatgcggta tttcgttttc 45840

tgttcttgca ttagtttgct aagcataata tcctccagct tcatocatgt tcccacaaaa 45900
gacatgatct cattcttttt atgattgcat agtatcccat gatgtatatg tactacattt 45960
gcttcatcca atctgtcatt gatgggcatt taggttgatt ctatgtcttt gctcttgatga 46020
atagtactgc aatgaacatt tgtatgcatg tgtctttatg gtagaatgat ttatatttct 46080
ctgggtatat atccagtaat ggaaatgctg ggtcaaattg tagccctgct ttcattctctt 46140
tgaatggcaa tactgcattc cacaatgggt tacactccca ccaacagtgt ataattcttc 46200
ccttttctct gcaacgtcac cagcatctgt tttttttga ctttttagta atagccattc 46260
tgatggatgt gagatgggat ttactgtga ttttgatttg ttttctcta ataatcagt 46320
atattgagct ttatttcatg tgcttggttg ctgcatgtgt cttcttttga aaagtgtctg 46380
ttcatgtctt ttgccactt tttaatgggg ttgttttct cttgtaaatt tgtttaagtt 46440
tcttataagt gctagatatt aaaactttgt caaatgagag ttgaaatat tttctcctat 46500
tctgtagggt gtctttttac tctgttgata gtttcttttg cgtgcagaaa ctcttaagtt 46560
taattagatc ccacttgta attgttgatt ttgtgtgat tccttttggt gtctttatca 46620
tgaaatcttt gtcccttcct gtatccacga tggattact tacgttgtct tccgggtttt 46680
tacagttttg ggttttacat ttaaactctt aatccatctt gagtttattt ttgtatgtgg 46740
tgtaaggaag gggccagct tccatctctt gcctatggct tgccagttat ctcagcacta 46800
tttattgaat caggagtctt tccccatga cttgttttct ctctctctct gtcttctttt 46860
ttttttttgt agatggagtc tactctgtc acccaggctg gcgtgcagtg gcatgatctt 46920
ggcgcactgc aacctccgc tcctgggttc aagcaattct ctgtctcagt cccccagta 46980
gctgggatta caggcaccca ctaccacgcc tggctaattt ttgtattttt agtacagacg 47040
gggtttaacc atcttgcca ggctggctct gaactcctaa cctcgatca gctttgtga 47100
agattagatc atgtgttatt ttcattacta tctcattatg atcaacaat ctataatttt 47160
tattatgatt tctttttggg caactgttac ttagaagtgt gctttttaat tttcaaaagt 47220
accaggcttg tctagttata tttttgttat cattttctag cttactttcc tgtatgtttt 47280
agtttttttg aatcagatga aatgacatat ggtctaatat ataataatct gatactcttc 47340
tatgtaccct tgaatacatg tggtcagccg atttgagaag tattagtttg tatgtttaca 47400
ttaggtcaag tttgtttgct gtgtagttaa aatctgctct atccttgctg attattatta 47460
tgtattccat ctttaataa ataagagaac ttgctaaaat ctgaagttct ggtgggtgat 47520
ttatcttttt ccccatagt tctattattt tcttacttta tttattctaa tgggcttaga 47580
attgctatat cttcctgggt aatttaacta gtattcaata tagagtatct ctctttatct 47640
ctagtaatgc ttttagcatt aaaaagattt ttatcttaca ttgatatcac tacaggattt 47700

ttcttaatta atatttgtct ggtataccat tttattgttg ttttaaaatt ttatgttttt 47760
atgttctagg tgcttccttt ttcttctttt ttgagagcaa gtctacatgt gcctaacctt 47820
gtgtctggta gcagtgggcc tcaggataaa aggggcagga agagcagagg aaatgggaag 47880
gacccaatt ttgctgggtg actttctggc acctagcagg gattcaatag atgtccattg 47940
aactgtcaaa atgaatgcag atggctggac ttccttctag tgttaccctt ggactcagat 48000
gaagtctccc cattccttct tctgtttga aacactagtt gtcccattgg gaatccctca 48060
tccccaaggc cttctgctca tgatgaataa tcaatcaggt ttgtgattca aaacaacatc 48120
catcatctct gtaaaaccag ttcccatcc taaagctatg acattgtttg acaagggtat 48180
gtattattta aaaccaggac tttgtcatgt tttcattatc tgagtacagt tggccctttg 48240
tatccatggg ttctgcattt atgaattcaa ccaattgtga attgaaaata ttcagaaaaa 48300
aatgcatcta tactgaacat gttcacactt tctccttgtc ataattttct aaacaatata 48360
gtataacaac aatttatata gaacttacat tgtattagga attaaaagta atatagagat 48420
aatttaagct tggttaatat atgcaaacac agcatcattt tatttcaggg attttggtat 48480
ccatggaagg ctatccatgg aaaccaataa ccccatatat acaaaggggac aactatattg 48540
gcatccttta ctgattatag aacattccta gtcactacca tagtgcctct ttcacattct 48600
ctttattatc ttcttctgga actccagtgg gatttttggt ggactttctc atcctattca 48660
ttatgcctct tcacttctct ttggtctatc tctttgtctc tttctattgc tttttgagta 48720
atttatttgg tccaatcatc aagtttacta tttctctttt tattggttcc aaccttataa 48780
tcaacatatt cattgagatt ctaactttat tacaattatt ttttcatttc taaaagttct 48840
actttgttct taaatttato tagtcaatta gagaattcct ctttcatcaa acttgcaatt 48900
ttctgcttaa tttctttcaa tatattatct attaatttat ttatttagag atagagtttt 48960
gctctgtcac cagggtgaa gtgcactggc atgatcacag ctactgcag acttgaactc 49020
ctaggctcaa gccatcctcc catctcagcc tcccaagtag ctggaactat aggtgtgcac 49080
caccaagccc agctaatttt ttttctgga gatagggtct cactctgttg ccagggtgg 49140
tcttaaaactc ctggtcacia acagtctcc tgcttgggc tcccaaattg ctgggattac 49200
aggtgtgagc caccatctcc ggcctatttt taaatattat gtaatactgt ttatctgata 49260
ttcccaatat atgcaatctt tgcacatctg atttctgtgt gtacacagggt gttcatgttt 49320
atacatttga gttttccaaa agccatagta cactcatgag aaaatgaaag tgaaaagaca 49380
aatacagtct tagtattaat tataaaaaata gtttttgccct aatggattca gagtcttagt 49440
gacctcaga ggtcctcaga ctacacactc agaaccactt cactaaagca tccatcgata 49500

caatcagtta tttctttcct gtgcctccaa aatgtttcta cctcaatacc aaccttggaa 49560
aaattgagaa aatgaaagct caaattatct tccatggagc ttaattttct ctgcaattct 49620
gatcaaaaaa gattgtaagc aatgaaagaa tcaaactagc tgtattcagt agtcctccct 49680
tatccatagt ttcactttcc atggtttcag ttaatagtc aatagagtaa gatattttga 49740
gagagaatga ccccatattc acataacttt tattacagca tatatttatc attatcatat 49800
ggatcagta tattattggt gttcatttct ttatgtgcct gatttataaa ttaaatttta 49860
tcatagttat gtatgtatgt ataaaagaaa acatgggtata tacagagttc agtactatct 49920
gcagttggaa ctcttcaaac atatccctca cagataaggg gagacctgct gtaatatttt 49980
tgcttcttaa ccaagtgttc tccctccacc catccctgtc ttagcatttc tatcattaca 50040
taattctcta tttggaaaac tgtatcatga ggtttttact gccacaatgc aggttaatta 50100
ggaaatgaga aaaacaaagt agcctatttc ttggtcttga gtttcaagtg atttagaagg 50160
aaaaatgaaa accaggctac ctggatccaa gaatgggaaa gagtgaggag tatgcaggga 50220
ggaggagaca agagcttaaa catggaggca tttctggctg ggtgcggtga ctcttgctg 50280
taatcccagc actttgggag gccgaggcag gtggatcacg aggtcaagag atcgagacca 50340
ttctggccaa cgtgggtaaa ccccggtgtc attaaaaata caaaaattag ctgtgcatgg 50400
tggaacgcac ctgcagtcct agctactcag gagactgagg caggggagtc acttgaacct 50460
gggagggtgga ggttgacagt agccgagact gcgccactgc actccagcct ggcgacagag 50520
caagactcca ccaaaaaaac aaaaaaaaaa aaaaaaaga ggcatttcta agagagagac 50580
cttaacccta aggcagacac tgaagaggat ttggggcatg aaataggagc tgcaggttgg 50640
agaacatcac aagcaagggc atctgtaacc caccctaccc acttttctca gagacagtcc 50700
agagagcgag taaggcacag gggccctctt ctgcatattg aaagtgagaa gaaatgccta 50760
tgcgagtgtc taggcagaga gcttgaaaca gccacttgtg ggtctctgca cctgcagcct 50820
ggtgcagggg tcagaagaat gtacttatcc tgtgcaacaa gagccatgcc tttgtgacaa 50880
agcccgtggc ccaagaaggc cctgcatttg gaacaaggag gcaccacttt gatctgctga 50940
cagctaagtg caaggtggaa tcagagatat cttaatagat gtcaatgata acagatgata 51000
ccaagaacca ggcattcttag acacacacac acatacacac acacacacac gtgcacgagc 51060
accacgcac gcatgtgact ggataccgca tgagttttct aaagcttaaa gatgactaca 51120
aggacaaaga gtaaacactt aattgactgc aattaagttt ttgatatcca gcagagtagg 51180
agtttttact agcaattaac ttcagtttta gaacacgaca aatcttattt ttattataca 51240
actacaaaca aatatataat gaatgctcag attccaggac cctataacctg gggatgatggg 51300
tgggacgagt taaagaaggg atctggacac tgctgcctct ctgccttctt tatatgggga 51360

ccgtatcaca aatgaggctc actgtgcctc cagtcattgtt tactgtgcag caaggctttt 51420
ggccagagaa cataattgac atgcacagag ctggagctct tgccagctca cagcaaaagg 51480
gcctctaatt cagagaaaaa ggagttggac ctctgccagt catgtacctg ttatgctttc 51540
atatagggcat cagttagacc ctgatctcag tctgacaaac cagaaaatat aaaaccatca 51600
ggatcaaggc atacttttta ctgggagtgt tagcctggta atggagagga atgttattgc 51660
tctaagcatt cgcataatat ctattgatca aacctctgtg aagctgctat tcttcatgcc 51720
agcgtctggc ggcaaggagg gactatttca ggtctcacca tcacagccac ttcacactca 51780
ggccatgcc aagtcaccag cacagcttgg gattggggca ggaagtcccc agggtgacca 51840
gacagaaagc agcctcagga cggggccaga gcatgagact aggggtggtg gcagctgggt 51900
tttatttact ctgggctgtg tgaccttggg caagctgcc agccattctg ttcctctctg 51960
catgccatac aaccgggaat tcctacatta ctgagtcact cctgggcca ggtttgatgt 52020
ccctattaaa atgctgttga gagagtaagt ggcttcagtc actagccctg gagaatgagt 52080
tcaccagttt gatttgttta caactgagaa ctctttcagt ttgtgtgctg gaggtgctgg 52140
gggagcagg ggaagaggct ttgctggagt gggggaaata tcctgcctgg gagtgggatt 52200
ggcggaagg ggcataattg tcaatcacia tagaatccta gacacttatg tctgggaggc 52260
atagcatagg ccattgcatt catcttcccc ccagtcactg aatctgtgac caatggaact 52320
gaggatcttt acttaacaac ctctagtgtg gggataatca cctccttcta aggcaagcta 52380
cttggaact ctaagatatt actgagaaaa gaccctgagc ccagccttga attagtccaa 52440
agtctaattg ttctcccagg tgacagccct tgaaagagtg tgtgacaggc gccatgcccc 52500
ctgaatggtc tcttctccca aataaataat acctgattct ccctctgcaa tgttcatgac 52560
ccaaacaaag ctgagcctct gtgtggccac tctcagtttg atgttgtacc ccaaaccttc 52620
aacctcagtc ttaatgcctg ggaatggggg aatgtgtgga aaaaggcatg aaacacagta 52680
cacagcaaaa tgtcttaaca tttctcttga ttctctagct ccttatccct tttcctaaaa 52740
atctctttct aaatctttca aggataaagg gaagggggta gaaaggggaa gtgaggagag 52800
gtaaaggaag atgttcatta ttgggagcct ataatgttcc aaacacgtgg gacatttaat 52860
cttcacaaag taagtatacc caccctaatt ttataaggta gaaactaaga cccaagatat 52920
ctaagtaagt tgcccaaagt caciaagcaa ttacctggca aacaatggat tcagataatg 52980
actatatcat aggccaagcc caattacact caaaataatt cacatcttcc taagaggcaa 53040
gcctggggg cccaagagg ggttggtgag taagggcaga aagtacaagg gagacaagtt 53100
gccaatcaca gcttggccct gattattgta ataactgaca caattaaagt tcaatttcgt 53160

atccttagaa ggtttccgtc atattccaag gctgatgaga aatttccatt ctaagaatgg 53220
ggatctgttg aatgttttaa attcaggcac ttaattaaat cttttaagat gactctttct 53280
gtgtttgccc ttagtctttt atttctatcc aatattgttc ctggaccaa tagggtcggg 53340
ctgctgtttc ttgtagccca ataatgagat gcagatgaac tggggaggaa gagagttttt 53400
tttttaagta tataaaaaca tttattcatt agaaaacaag gagactggca aacatatatt 53460
ccaaagtga agcagctcaa tgcagttcag ttaggctaata ttaagagaaa ggcttgcat 53520
tttaaagatc gtgtatgtat tttttttttt tttcaaaaaa ggagacaggc aaatattcta 53580
caaggggaac agaattagaa ttctaggcca ccctacaagt taccctgcac agggaggaaa 53640
ggaacaggca agatgacttc tcaggatctg tgctgcgag ctgatgctct gagaatgggg 53700
gttattttct tgggtgtcct gtcttctgtc atctaggcta aaaaatcttc ctacttgac 53760
tcatcacttg agaagacaac ttttggtttc ttttcggaag ctgctgctg ggtgttcgga 53820
tgccgatggg tagtacgacg gggctctggg gtgacgaagt cattgtctga ggctgcagaa 53880
gccagaggct ggtgcctaca aacagtggat ggtttcttgg ctgttggggc tctctggcta 53940
cctgcttggg taatctcaag ctctcaatc ccctccaaac ctctggcatg acaggcctga 54000
agcttttgct caaattcatc tattatagcc ttgccctcag gcttggcccc atgttgtagc 54060
ttgcccataa ctgacaaaaa agaactgaag atggactcat ctgacagttt atctccaaaa 54120
cagtcaaaat tgccaagcat cttgcaaggg ccttgctctg tgaggggcag ctgtgacaca 54180
cagtagggca ggtcctgctg cacctgctca gtttaggctg tgtggaaccg ctgacacagc 54240
ttttccacca ggctttgtct gcttgtcttt ggtgatgatg taggagaaga actgcttcat 54300
aatggtgtga aaaggcactt cctccacccc cagctcgggg tctgacaggt ggctgatgat 54360
atctggaagg agactataga ttgcattgcc cttgtggaag agctcattga agaacttctt 54420
ggccagggca gcaatttgag actcagggtt gatgagcagc atggccatct cactcacctg 54480
cccctttacc ttcacatgt ccttgaggat caggtggagt catcaccagc cccactgttt 54540
tccacactgg ctgagcaggg tcctgaaggt gagcatatag atgaggagtc caggggtcca 54600
ccagattggg aaagtggatg gccagatccc cagtggcaac gatgagatta gaccggacaa 54660
tggaagtgg agacttttct agcatggtga acagaagatg aggctgggag ttgcagaaag 54720
tggtactgat catgcagaac ttgccaaggg taggtgaagc agctgcagag aggtctgggt 54780
tgctatagag gctgagttg ttgtagactt taagcaagag tggaacaaag gcagccagtg 54840
tctgtttgcc atccaacagt tccatcttgc agaagccgcg gattagttct gctccgtgt 54900
catctgctgc cccaaccagc ccagctcct cctccatagt ggtctcgaaa cttgtattct 54960
tctccttggg atctttgggc ttgtgctcct gttcttcccg gagaactcgg cgctgacaga 55020

gctctccact cactgcctgc tccaagtgga ccagctgctg cagagccaca tccccagcca 55080
gggacaagag gttoatcaac aggaaagtgg ggagcattgt gggagactcc ttcgggtccc 55140
cctgactggg tctcttctct tctagcttct ccagggcctg ttttgacag ccctgccata 55200
tctgggcaca gatcactgtg ggactctctg ccagttggta aatgaggggc actgccacct 55260
ctttgaatgg gatccagagt gggctctggg gacaaagcct ttcgggaccg tctcccgag 55320
tcactcaaac aacttgtgtt cctgaggcaa ctggaagtgg gggtgacgtt tgcccagaga 55380
agtctttcta ctgtcagaga tgttggcgat ggcatggcac acctgctggg ccagccggtg 55440
gtcctgtgga aacttcggaa gagagtttta ttttctgcaa ccggtgacgg ggagaaggcc 55500
tggaattat tgccagacca acttaaaatt acaaagtttt ccagagctta tataccttcc 55560
aaactatatg tctacgtgta aggtatgcat tcatctaaag atgtaaatgg ttaacttctt 55620
ttaatctata accaaggtct gagtcctaaa taccttctc tggagcctca gttaaatttac 55680
ttaatctaaa tgggtccagg tgctgggggg attaccctta tcttgtctcc tgctaaatta 55740
cagaggtttt gggagttcct tcagacctcc aataaacttg tttgtggagg cctaaggagt 55800
ttccttagac cccagtgaa acttgtttaa tcctaaatgg gtctgttaa gaattccttt 55860
gttattttgt catgccttaa gtcccaggaa aggcctaggt aaaactcttg atgggctttt 55920
gttacattcc agccttcgta caggggcact ggcttttaat atttaactta accactcagt 55980
cagtactgaa agagtgtca gtgacacctg gcctgccaca attatgagtg tccagaattt 56040
tactagatgt ttgggggaag gggaatgaga agcctgaatt aggaatattc tccacgagtc 56100
tctcacaatc tagttggaaa gactagtaaa ctcaatgaca tattaatga taagtaaatg 56160
acaacctgtg gtccttactt catgccttg cacacactgc tctctctacc tggatgccct 56220
tctctcacct gatctctcca gcaaagcact actcattcct caagacacaa agctgaacag 56280
gcactctgtg aagcctgcct agtccgcttg cttccttccc agaagatttt gccaatcatc 56340
cctcttttgg acaaatatcg tgcaaacactg tattttaatt atttcccaac tccctgaaac 56400
actagaatgt gagcaatttg aagataggaa tggagtttcc ttccttgcat tctcagcaca 56460
tagcattgtg tctgaaacca aatagtcagc aatcaatagt aattgatgag tgtttaaaga 56520
agcactacag gaagtggcaa aaatcagggg cctccagggc aacacgatgg ggacttcaca 56580
gagtgggtgg agcttgattt gggccttaga ggatgactaa catttgaga tgctgatgct 56640
gccaaaagta gaaagatcag gaggaataaa agcccagaaa tacgagactg tcttccccag 56700
tgttttatgg tggtcctaga ttcaaatac ccatcatgtc acaactttga aatgaggaaa 56760
atactttcac caagaagttg aatgctattc tgaattacaa acccaggtga gtgctcatca 56820

agaacagaag accagccagg tgtggtggtg cacgcctgta gtcccagcta ttcaggaggc 56880
tgaggtggga gtatcacttg agcccaggag ctcatagtgc gctatgtgga tcggatgttt 56940
gtactaagtt tggcatcagt atggttacct ctatcccagg agcagggtgac cactagattg 57000
cctaaggaag catgaaccag cccaagttgg aaacagagca agtcaaaatt cccatgctga 57060
tcagcagtgg gattatgcc a gtgaacagcc actgcactac aatgtgagca acataaagag 57120
acgcccctgt ccaaaaaaac aaagaaggaa acctgcaatc cgagcacttt gggagggtgg 57180
ggaaggagga ttgcttgaag ccaggagtcc aagaccagcc agggcaacat ggcgagaccc 57240
tgtctctaca aaaataaaaa taagttaaac ataaataaaa aattctaaat gaattttaaa 57300
aagaaagaaa aaataataga aggctttggg gatcagggca gcccaaaggg attcctggag 57360
tttcaaacac aaaatagaca tgagagactt aatctggatg tcagaccagc actagaagaa 57420
aacctggaca ttttgaagcc cttaaattca taaaactttg acagtttcat gaacatttgt 57480
tgaaaccttg ccatgtccaa gaaggtcacc taggatcaga acttccaccc atctcctctt 57540
ctccccctgt ccttttctga ggtctacacc cctttttag taggtggtca actggacaca 57600
ggactaggtt tatcaaaact tgctctcggg aaataaaaat aagacaagaa atacaatata 57660
cacaagaaca catgttcttg tacctcagtc tctgaatagc ctgtctctag actctgtttg 57720
atgtaagagt ggttgcttct gaaaaaagaa ctgagtgact aacaatggtt ggagaaggg 57780
gcctttctta attttagttt ttattatttt tattgaagta tgaagttcaa tccggtgcac 57840
gcaagtatca gctgatgaat tttcaciaac agaagatacc agtgtaacta gcacccaaat 57900
caagaaatgt taccatctgc ctctatgccc cctcctacca attcccagta ccacctcttc 57960
ccccaccaa gagtaagccc tgaggcccag acaggcttca cctaatttta tattttattt 58020
tttcttttta tgaacagctt tattaataa ttcacataac aatacagttc acccattaaa 58080
agtataaaac tgagtggctt ttaatataat cacagatatg tataaccatc accaaagtca 58140
attttaaaat attttcatca cctcaaaaag aaggcaaata ccttttagct ataactccca 58200
cctcacatcc tccgtaaccc tagccaactg ccaatctact tcttatgtct atgaaattgc 58260
cagttctcca catttcatat aaatggaatc atgtaatata ttatcttttg tgactgggtt 58320
atttcactta acactttcaa ggttcatttt tggtgtagca cgtatccata ttccattttt 58380
tatggctaaa taatattcta tcatatggat atagcacatt ttgtttatcc gttcatcagt 58440
tgatgattat ttcagtgggt ttacctttt ggctcttagg aataaaatgc tgctataaac 58500
attcatgtac aactttctgt gtggacagtt ttttctcttg ggtatgtaac tagaagcaga 58560
attgctgggt catatggtaa cactgtgggt agcattctaa gcaactacag ttttatattc 58620
ccaccagcag tgtacaaagg tttcaattcc tctcatcct caccaacatt tggtgttttg 58680

ggattgttca ttacaaatac atgaaacagt ggttaccaga gtcagtcagg gtgggaggga 58740
ggaatgcaga gatttaggtc aaggatacaa agtagcaa atcgtagaatg atttaattct 58800
ttacataaat agaaccttac agtatgtgtg cttttgtttc tgacttcttt tgctcagcat 58860
tagattcagg agattcatca acattgctgc gtatagttgc agacagttta ctatccttat 58920
tgaatgggtgt tctactgtgt gaacatacta cagttgcttt tttcattcca ctagtgggtg 58980
gcatttgggt catttcaggt ttagggctac tatgaatact gctgccttga acagtccagg 59040
atacatcatt ttgtgagctg cttatcttta ttatttgtgc tacctgaatt ttgtattatg 59100
tgcaggattt acttttttta aatgtcctaa aaacaccttt aagggtcttt aaactaggat 59160
cagaaactta gctccttccc tactgacaag caggcagaaa aaaaaataa tagtaatagc 59220
tagtatttac tgaatgctca ttgagccacc tctttaggga aaaaaatctt ttcatatatg 59280
tttgctttta atcctccaaa ttactcaaag ccataggtat tctcattatg cccatgttag 59340
acctgcagaa actgagaccg aaggagggtta aataacttgt gcaaaacct acaactagca 59400
agtagggaat cacagtacaa atacaggtct ctgaatataa agcctgtgca ttcacctcc 59460
aattacacag cctctgatta cactcagctg ggagttcagg gagcctgcta gacatgggga 59520
ggagaagctg aatgagagc acctgaaact gccagctcag tccagctcaa cagctcggat 59580
tggtctcccc accgcccccc tacttcctgt cctaattaat caaatgctct tgaacatgaa 59640
agagagttag attggatttt tgtaagatg aaagaaagga atctgttccc atctggcacc 59700
ttcaaaactc tcctattcaa gccagcaggg agcatttatg caacctacat ggcagtttaa 59760
taattcccca agaatgccag gagattaagc aatgtcaca ccatctgga tccaccggca 59820
cttgagaata atggaaagag caggggtcta aggtctagga caggggtcaa ccaaatttct 59880
ctgtaaaggc ccagatatta aatatttttag gctttgcagg tcctatgatc tctgtgtcaa 59940
ctaccaact ctggcattgt aacatgaaag cagccatgag caaacaggca tggctgtgta 60000
actaaaactt cttttatggc acagatattt gaatttcata taattttcat gtgcctgggc 60060
atatatttct tttttaaaaa atattttcta accagtaaaa tatgtaaaag ttattcttag 60120
tttgaaagct gcacaaaaat ggaggacaaa gtctagcgac ctgaattcaa atcctctctc 60180
tgccatttgt tatttctgtg gctttcgaca tattacttac attcactaag cttcatttct 60240
ctotttgaaa aatgagcaga gcagggtcca tgcattcagg gcctcccttg tggctggcac 60300
tgagccacat gctttatatg ctttactgaa ttaaatccta caataaccac aaaaggctga 60360
aattatatga ttgccttctt cagatagtta cgatgattaa gaaattatgt ctataattgt 60420
gcctgagctg gcctocatag ttcattgagt ttgatgcttc ctgttgtacg caaaaattgc 60480

catgttggtt cgatgagagc atgcctgctt tttctactac ctgtgggtgtt ttgagtagcct 60540
atatttgtat tgcacttaca gtgaaaatta tctaattata ttattctatt taattacaat 60600
tgcacatcaagc tccaataaac aatccaaatt tcaactgaact gaaagttgag tgcacacttt 60660
gaatagtcac tccttctgtt aggttgggtgc aaaagtaatt gccattactt ttagtagcaa 60720
aaacagcaat tactttcgca ccaacctaat atatatgtgc acagtaggca gtttgggtca 60780
ggtacactta ttcaaaacat tgcattaagt aagataatac atgttaaata cctagcacac 60840
agcctgggat gcaggtgtcg gccagtgttt gttccccact tccttagtga caaagattag 60900
gcttagctac tttttgggaa aggtgagtgc caccagaaag atgaggccaa ggccttatca 60960
gacacactgg aacacagcct tccaaggaa taaatccaga atattttgtg tatgagttct 61020
ggctctgaag ccagcctgt tcttactctg acccccagt ttctgccttg atcagccacc 61080
cactaggctg ggcattgctt cctctcactc tgctatctag tgaccaacc tcggccaac 61140
cctccagctt tcttctggcc atgcgactgt tctgctagtc tgctccgcat atctgctcct 61200
accttccgat gcctcctctg agcttgggt ccactccca ctgagcaca acaacttact 61260
ttgccccagt atgggtgaca acagctcaga ctcttctaga ggtctaatac tggacacctc 61320
gtttgcttta taagcacaat ctggaaagaa gagtttcatt gctcagagcc cctcaggaag 61380
cagaagagga gagtggtgaa gagagcattc atcctctgga atgaccataa ctgggtgtttc 61440
aatcccagct cccctccata ctaatacttg ggtgagttac tcaccaggctc tatgcttcag 61500
tttcttcac cgtaaaaatgg aggttgtaaa aatatcatct accccatagg gttatgggtga 61560
tgattaaatt agttcttcta tagaaagtgc ctgcagcagt ggtagcatg tggtaggcac 61620
tttgtagtg ttatctgtta tgtctgtttg ctcttactaa catctatatt tctgagtcct 61680
agacctcccc agagagtgtg aagttcatgg tgtttcatgg gaacagatgt aagcccactt 61740
gctgaatctc tctctgagct aataattaac aaatacattg cacagtagct taccaggaaa 61800
gcctatactg gtctgaacaa gaagaaactt gctgaacaat tctaaagctc acttctcaat 61860
ataacagcag cccttgctca acatgttgcc catatcttca gtaccgtgca gaggtcacag 61920
ccatcatatt ctgccagctc tccccctatg ggttgagcaa ccataatgtgt cttcccaacc 61980
aggacathtt tcatttttag cttcttggtt tgacataatt tcaggtttac agaaaatctt 62040
caagagtagt acaaatcccc ccatgtaccc tgtacacata ttcctgaaat gttaatattt 62100
ttcatatttg ctttatcatt tcctctatca gtctatctat aatacctacc tatctaccta 62160
tccatccatc taaaaatgtc ccaggtaatc agcaaacctg ctttaaagt caccaattat 62220
cccaagaaca tccttttttag cacaaaaaac cccgaatcat gtattgtatt cagttgcat 62280
gtacgtttag tgttctttaa cttagaacag ttccttggtc tttctttgta tcttatgaca 62340

tcaacat tttt taaatagcat aggttagata ttttgtggag tttctctcac tttgggtttg 62400
tgtgatgttt ccttctgaat agattaaggt tatgtgcttt tgagagttat acaccaggaa 62460
tgatgctatg ttcctcttag tatatagtat cagcagacac agtatatcaa tttgtcccat 62520
tactagtgat gttattgtca gatttctcca ctgtaaagat actat ttttc cttttgcaac 62580
tgacaagtat ttttaagagta catagccatc ctgatactcc ttaaactttc actcacagca 62640
cttaccatac agtgtagatt cctgtctgaa tcaatgctat aaagggtttcc aaatgatgat 62700
tttctaattt cattatttct tctatactta ttagcagtct ctctacttta aggaagaaat 62760
ttcccttctc tcccattttt aaatttacat cataggttct tcttttattc tgtgggttat 62820
aatttggtgc catagttatt tctttttatg cctagattgc tgactggggc cagcaagagc 62880
cctacaattt ggtgcgtgtt tctgttgaca tgtttctcag catttttggg catgtcctca 62940
cttgccagca caacatgatg ttccaggctt atcttgtcct tttcccagcc ccaggccagg 63000
aatcagccat ctccaatgag ccaagtaaaa ccatggcata ttttaagagaa agggccacag 63060
gcataatcat ggcgggtata aactgaaact gacctagaaa tgctgggtca tattcccaat 63120
ctaccactag atgtccttcc acacctcaat agttcttcac aggaagcagt cctggggcat 63180
ttggaaatgg tgggatattt ggagttatca caaagactgg ggcagcaggg gttacttctg 63240
gcatttagta agcacagccc aggcattgcca aatttcttgc acaggaccat ctatcctaac 63300
acagaattgc ccataagta gcaatcctgt taataaatag tggccaatc cctgaactaa 63360
tctatcttgt gttctagtct ctcaaggatt ggcttagttc ttttcacott atccgccatc 63420
agttaattta atcaccactg tttccatagt gagaatctaa tatattagac tgatcaagtg 63480
tgtaaatcaa atatattgga ttagctggca tgcattatag caaataaaat agccagtctt 63540
cattttagat ctagttaaa gaaagcctct acttcttatg tgtcaatttt tgcagaaaac 63600
attattctag caagttggca aggagtctag gacttatgtt aagacacacc atctcatccc 63660
cagttcttgc agctctctgt ctagtattg gccctattca ttgcatcagc accaacaaca 63720
gctaccaggg ttttataggg caagcatttt actaaacatt ttacataaac tatctcattt 63780
aaccttcaca gtagccctat aaatcaggca ttcttattac aattttacaa agagccctgg 63840
attccgtgtg gtcagtagca agccaagat agcacagctg tcatgtagca gagccaacat 63900
tcaaaatccc acagtcgggc tccagtagcc atgctgtgta ctgcccattg atgtaagaac 63960
ataccaaca caatgaaaaa taagaaatga aatagaaagg gcatggaaaa tatcaaatca 64020
taactgcaga ggaaaaaagt caaatttaag cttagagtta ctgcaaagtt actgcaaccc 64080
atcatatttt attttgtgct tctagtttag tggctaaaac taaggggcaa actgactaaa 64140

tataatattgt tgtaatcttg agggaaatgc acagcaagat gtttcttgat caagaaaact 64200
gatactacgg agctttatat tcagtacttt gatgaaacag cactaaaatg tcctcaataa 64260
catctctaca aagaatcagt gattcttaga cctctaagag gagccagtgt gatgcctaaa 64320
cataggcatg taggctactt gacagacaga ctcagcatag cagtcatcat cacgtagtaa 64380
tttctagcat cctagtagct ttctcactgg ctttagtttt gaataaatgg gatttttgct 64440
aagatcttac attatgcaac agaggcaacc acaaagacca cccctagaaa atgaaatgat 64500
cttgggcccc ttctccctt acctcctgct tcacactgca ttctttggga agcggatgct 64560
gaaacagagt caggtataca aaaatctggg gagaagtcac tcctgtgaaa gcaaagggaa 64620
gaagtaagat tgggctggag gaoccatcag agcctgacat taacctgaca gtctctggca 64680
gctcaggaat cctttaggat ttccatgttg ggtataaatg attagactct tctactataa 64740
cttagtcatt ggtagagggt gtaccaagaa tagcatgatc ccagctaâaa agctgaggca 64800
aaccctgagg aatctaacag ctggaggatg tcagcaaacc atatacctca cagctggaca 64860
gcaaggtctt tcttgaaggg ggatctgagc agcatatctc tatgtctgcc tcatttccct 64920
atgtacaagc gtgtacctga ttgcagcaga atgagatgag aaacaaatat ttttttgtaa 64980
caatttacca cttccatctc agcctgctac cccttagctc tccttgggca aagggaaaaa 65040
gacatttgaa gagcctgcct cttggataag tgaaaatgga actctcctct ccttctcttg 65100
gatttctaaa catcccagat catttctgtg aattcttttt attaggaaaa tagcatcctt 65160
taaaagagaa aacaagttgt ctatgtctta tattcagagt cagtttgagc cccatgttgt 65220
gcaatgatct tgtgcaatga taattacatg ctaccatgcc tcaagaaatt tccccttaag 65280
tggagtgagg aggagaaata gtggaggctc aatgtcttgt tttaagaaaa cacctgttat 65340
tgctgaacct gatgcttggt cttatgagca caggagtttt ttgtggccca gcaataacca 65400
acaactcagg taggccagtt gcctcattta atgagcttct tggtttctga atacaacaga 65460
ggcaaagaga gtaagtttta cgaggacttg ggtttcctta attagacaca agctttttat 65520
ttggttgggg gaactaatgg agaaaggag acactagcgg catttgtaac ttgttaagac 65580
taattgaaaa tcttaggacc aaagctttct gtgaagagag tgctctaata agaataaatg 65640
tgatgattat ttacctaat gagcagtaat gactgaggtg tgagccactt cccaggatt 65700
aataaagagc cggggagtgt acagttaga ccattaggct tagctggtac ttcactcttg 65760
cgaaggacac agtgccagaa tcatattaat attgtaaaat gagcccaaaa catagacata 65820
tacttttgga gaaataatth taattggcta agtacctgac atcagggttg gccattcagt 65880
tttatcatta aacaaagttt cacacaagtt cttcaagact tgtgccattt acagtgttct 65940
cagcattcag tgttttaaaa agaaaagtag aagagggacc ttatcaata tcaaaaatag 66000

ttttttgtgc aacaaaaagc aagcaagcta ttattatata tacacactcg ctttaaaaaac 66060
aacactaaaag acattgagtt gagctggaag cattgcatac tctgagatac acagttgata 66120
ttgttaccca cagtgtccct tcaggaaagc cagaggtcat ctgcaacacc acaattaagg 66180
ataaaaggac cttatattat cagaaatttc aatgtcttca taaaaattgt gtttatattc 66240
aatatctccc ttatttagtg tgacaaatgt tcatgtacga gtcactctca gaaaaagagt 66300
tactcagctt gctacgtatt cattattcta tggctctgct atgcagggga aagtgc aaac 66360
agcaactgag agaaagcagg aatctaagac atgggactca acttgccctt tacacagggg 66420
gaatcctaca ctagtgacct ttgtgtttta aggaagtttg caaatatggt taattggaca 66480
acatttcttc ctttatctag caagcactga gtatcttgcc aggcactatg aaaaaagata 66540
aaactataaa catgagcaag gcacagccac tgccaatcag aagcttaca tctagcagag 66600
atcttaaaat gcaggagct cttttagaaa tctgcttagg gatcaaaatt tccaatcaca 66660
cttttccttc cagcatcttt attctgtaat cattatttta atcttcattg aacgagatga 66720
ttaatgtggt tccatctgct gtatgaaatc acccatgcta cagtcagcca atcaagcaac 66780
agcagtaata atgtactact tcatttaatg gaaatttacc aagctgttac tatgtaatat 66840
cacagtttta gggcctggag aggtagagct gaactaaaaa taataatctc tgttatccca 66900
tagtatagtg gaataggaaa gaaatacaaa aaaaaataat gtgttcatgt attataggta 66960
ctataatagt agggtaggaa gatacacaga agtggaata accaagccaa tctcagtact 67020
cagaaaagggt gttataaagg aggtaacctt gaattaaacc ttgaggaaat ggatattctt 67080
caggtagagg tgtataaaga gcttcccaca ccaatggagc agcataaaga aatggcatgg 67140
cagagtgaac agttgggtat ggccaaggaa actctcaact gcatcttctt cactgcctca 67200
actgatcact caacgctctc cattctctca gcaaacaccc tgactaacta aactcccatg 67260
actagcagga tattattaga cttacaaaac ctgtcaattt atatgctttc cttttagaaa 67320
gacttgactt ttttttagtg aactgcaata gaactaagct caggagatgc aactatgaaa 67380
gatgaaggaa cattataaaa atctctgcaa tgagattctg ccttagattg ctttgctggt 67440
atgtcaactg attactccaa tttaaagaaa cccaaactgg aaattataaa agataaattg 67500
tgagtgtatt ttatgcaaaa aagaatcaaa tcaaatcaaa gtcttggtc tacattaaac 67560
aaacgacaaa ggcagtgttg tttttatcaa aatgttttag ggtagatttg tatgtatgcg 67620
ttatttcttt aaccaatggt ttaatcagtt aaacagctgc tattattaat gtgctggttg 67680
aaagtgtttt tottaaagta caaatttaaa atagaatgag aaaagatcag acttgagaag 67740
aaggaaagtc acagtaggac atttggaatt tattctctag accagaggtc tgcaaaactcc 67800

agcctacagg ccaaatccat cctgcagctt atttttataa ataaagtttt attggattac 67860
agccaggccc atccagttac agaattgtct atggttgctt ccactctaca atggcagatt 67920
tgagtagctg ttacagagac tatacgcaag tgtgcaggag agtttagacc acacctacag 67980
gaaaggatg aaatactacc ctccctccta gacacttccc caaacaaaaa acttaagtca 68040
gtgtgggaac aaacactgtt gccccacaga gcactaataa aggcagataa atgactgaga 68100
aaaccctatc ccttgggggac ggaggccaga aatggcaaca ccaataccat tggaagtctc 68160
ctacttctag gtaaggacat gaaattctct catatgcaac acccatcaca gatacaaggc 68220
aggagtttga ttttcatggg gaggaggaat tggaacactg agaatacccc acccctaagg 68280
cccaggcaca ctggcctata tgggtctgtg gctgaactgg gacaaaagag aagccctaaa 68340
accaccagta gcaagcattg agtaacaagc agtagcagta ttacagcaaa acaggcacag 68400
ctgaaagcag agagtggggc attaaaacc ttcagcgtct cagctactat gctaatagaa 68460
aatccaaaga aaaatctgaa gcccatagga ggctgaggat gaccctacta acaacaaaac 68520
tcaagtctag ctagactcct aagcagattg atttaatacc ccactctaac aaccttaaag 68580
aagtgtgccc atttccatgc ataaatacta tgtatctcaa tatttgatat actctacaca 68640
aaatgtttgc catttaatca aaatgcatga gactcaaaaa gaaaatgcaa cccactgtta 68700
acagccaaag caatcaacag aaccaaacag agacatgact ccaatgttgg agctattaga 68760
cacggacttt aaataaaata attacaatga ataagaaaaa tacattaaaa atacttcatg 68820
aaagaggata tatggatggc aaagaagcac ctttaaaagt gttcaacagt ttatacacta 68880
ttgatggaaa tgtaaattag ttcttccact gttggaagca gtctggagag ttctcaaaaa 68940
acttaactac catttgaatt agcaatccca ttactgggca tatacccaaa cagaagaaaa 69000
ctgttctccc aaaaggacgc atgcactcat atattcatca tagcactatg tatattcaca 69060
ataccaaaga catggaatca acctgaatgc caatcaatgg attggattaa aaaatggaat 69120
actatgcagt cgtagaaaag aatgaaatca tgtctttaca gcaacttgga tgcactctgga 69180
ggccattatc ctaagtgaat taatgcagaa acagaaaacc aaatgccaca tgttctcatt 69240
tataagtggg agttaaacct tggatacaca tggacataaa gatggcaaca gtagagggga 69300
aaaggaagaa ctagagcaag ggatgaaaat ccagctattg ggcattatgc tcagtacctg 69360
tgtaatggga tcaatcgtac ccaaattctc agcaccacac aatataccca tgtaataagc 69420
ctgcacatgt acccctgaa tctaaattaa agttgaaatc atttctaaaa aaagaaaaga 69480
aaaatgaaca aattgaaaat aaaaagatgt tcaacatgat tagtcattag agaaatgaaa 69540
aataaagcca tgatgaaata tcactacata tctagtagaa caactaaaag aaaacacact 69600
gacagtcttg agtgctagaa agtgtgtgca accacagtct ctctcataca ttgttggtgg 69660

gagtacacag tgggtgcagca acattggaaa actgctgggc agttttttat gaagttaaac 69720
atatattttac ttactctgtg gcccaacaat tccacttctg agtatttgtc ctagagaaat 69780
gaaaattttat attcacacaa aaacctgtac atgaatgttt atatcagcct tgtttgaaat 69840
aacaaaaaac tgaaaataac ccaaacatat ttcagtgagt gaataaacat attgtaatac 69900
atttatacaa tttaaaacta ttcggcaata agaacaaca aatgatacac acaccttaga 69960
agaatttcaa aggcattatg atgaattatt tttttttttt tgagatggag tctcactctg 70020
tttcccaggc tggagtgcag tggcgcaacc tcagctcact gcaagctccg cctcctgggt 70080
tcacgccatt ctctgcctc agcctcccg gtagctggga ttacaggcgc cagccaccac 70140
gcccggctaa tttttttgta ttttttagtag agacggggtt tctactgtgtt agccaggatg 70200
gtctcgatct cccgacctcg tgatccaccc gcctcggcct cccaaagtgc tgggattaca 70260
ggtgtgagcc accgtgcccg gcctatgatg aattttttta agctaatttt taaagggtgc 70320
atattttgta attccattta cgtaagattc ttaaaatcac aaaactataa tgataaagaa 70380
cacattagtg tttgctagga aatgggtggg tagggagtgt ggctataaag ggatagcaca 70440
aggaatttc tgcacagtgt tagaactgtt ctctatcccg actgtggtgt tagtgacatg 70500
accttatatg tatgtgagat taaatttcat tgatatatga catgcttgta aaaattggtg 70560
atgtgcaaat aattctatag tttagttaac agtattatac caatgtcaat ttcttggttt 70620
tcacattat actgatattg accaggtaga gcgtagatgg gaaacttctg tcttattttt 70680
gcaactttaa tatgggtctt aaatgatttc aaaataaaca ctttaaaaca ttctccctac 70740
taaataacca tgattaatag gttaaagttt ccaggggaaa aaagtggaca acatagatga 70800
acagttgagt gatttcacag aaagataaaa actataagca agattcaagt agaaatgtgt 70860
gggggaaata aatcacata gtaattcaca gatttaattc taggaatcgt agactttaca 70920
cagccagagg aagaattagg taatctgaag atagatccat aaaaattgcc caaagtgaaa 70980
atcagtaaga aaaggacaaa gaaagtggat gaggagaaca aagtatttaa aaatgatggg 71040
acaatatcaa atagtctaaa atatatgtca ttgaagttac agaaggagaa gagagaatgg 71100
gcaaaagaaa tctctgaatg atagctggaa ttctccaaa ataatgaaag catcaaatca 71160
tagatccaaa attcttaaat aacctgaagc agaataaatt ttaaaaagct agctataact 71220
tatccaaagt accaaataaa ggtaaagaag aaattgaagg aagacagtgg aaaatatatg 71280
aacatattaa atcagagaaa ttcttaaaga atacaaaatt acagctagat agaactagaa 71340
taagttctag tgttccatat tactgtagaa ttactatagt taacaataat atatggtttc 71400
acataactaa aagaatattg catgatcctc acaaaaaaat gataaatgct agagatgata 71460

gatatgccag ttatactctg ataactatat gtactgagac atcaccatgt agcccatga 71520
atatgtacac ttattttgtc atttttaaaaa ttaaattaat ataaagaggt aaagacaaga 71580
attatagctg acttcttctc tgaagcaatt aaagccagat gataaaacat atcttcaaag 71640
ccagctgggg ggacaactat caccagaat tctacatcta gcaaaaatac ctgttttaaaa 71700
tgaagataca ataaggacta ttcaaaattt ttgtaatggg agaattcctt accagcagac 71760
ctgtacttta agaaacatta aagaaagttc ttctgacaga ggaatgtgaa accagaccgg 71820
aaattttatc tacagagaga tgtgaagaac tctgaaaata gtaaaaataa agattaatat 71880
taaaagccac atttttctta atgttaaattg actctaaaat gtaattgact acttgaagaa 71940
aaattgtagc agtgagattt ataaccatg taaaatttaa atgtctgata acaatagcac 72000
aaagtttatg aggaatttga atgtctaagt cttcacaata tgtaagaatt tgtataatat 72060
tatataaaac atctaatact cagtctttac ttgctgcagg gaagatcaat ttggcaggtc 72120
cagcactgga attaccattt agattagtta ggaaactgcg gaagttcaag aaagaaatca 72180
tcagtaccta gagtatgcca aataatagta agaatacaga gaggattcaa atttgagatc 72240
tacttaataa ggtaaaaataa ggaggatgcg gtgactgagt attgagatgt tctgaatgtg 72300
aggtagaggg aagagtctac tatgtattat gggttcctgg cttgggtgtc cctagcagac 72360
agtggtgcta tcaagtgagt ctgaggagag aggaggatag gatcattaat atgaaaagtt 72420
aatgagtttg gattttaact ggggtttcag gcaatttaca tgtaatgtta ttattgatgt 72480
gtttggattg aaatctacca ctggtttctt gcttttattt gttccatcta tatttttctt 72540
ccttttttcc ttctatttca gcttgatta attgagcttt ttatgatttc attttgtctc 72600
cactattggc ttattaatta tgctctttt tctatttctt ttaatgttgt cctaggcttc 72660
acaatgtaca tcctctctac attcctacag ttatttaaca tacgtttctt ccttgaactt 72720
ctgcatagtt tcctaattaa atggtaattc cagtattgga cctaacaaat tgatcctcac 72780
cacagcaact aaagaaatct atttataata catgtttgga gacataaaca ttcaacaact 72840
ctcacagtat aaaaaatac atcctctttt aattggcatg tgagaccag aagtctggat 72900
tcttcttggt tctgcagacc catttttgcc acttcctagc tcctatttag gcttcagcaa 72960
taacaaacag atgcttaata tacacaagcc aagtagtttc tcatctccat aactttgttc 73020
atgctgttat atgtgcctag aatcaatatt ctgcattcct ttctcacctg accaatcctc 73080
actgaaaagc tcagctattg cccattctt gacttcttcc ttgtccacc gcttttgtgc 73140
tccaactgga acccattca tgtttacaag aaacactcaa aggacaataa agcctgttgt 73200
gactctactt cggcttttct gagtattaaa atttcctaag atctttcaca ggacaaattt 73260
ttaaagcaa gagtaataca tggaggtgat taagccttta aagtgtaccc tttgttccca 73320

aggctcttggt aattatcatt taggagcagt ttatctgcaa tgtgctcaat tacttaggta 73380
actaaatagc tttttactga tcagcattga acatcatttg cttaatagca actaaciaat 73440
gcggtggcag aattccactg atagtctact ttcagaaata ggcttattta agattgagct 73500
ccaggggaaa aatagattct aacagaatag attattgtta ttacataac cacttaggtc 73560
tagagaaaca aaataatctt tccgtttttt aaaattaagg aatcagtaag ccaaagaga 73620
atatccaagt aactccaag tccattccgt tcattcatcc agtccatttg tgccatgaaa 73680
atcttcatga cattcctatt tgatcaatca caactcacag ctctcaatga ctgctcacat 73740
ctgaccttct cagtaaggac ttttcttata gcttagaaca gtcataaggat ttttctctc 73800
caagcttttg acactttgga ttatagatcc catgtgcttg ccaatgcttt ttttggttaac 73860
tgtagaaaaa taaaagcaa atacagacag gcgaaagact cagaggggtca cataacattg 73920
tatggtgaac tataccatgt gaatggtcaa ggaaagcttc tggagcatgg gacatattag 73980
ctaggtgttg aaagaaaaat aggaattcaa taggctggaa aatgggggga gggattttca 74040
agcagagggga acacctatct agatgcacaa atgtatgatc aatatggtgt tcaacattga 74100
aatgccacat ggggacctga caagaggtag ttgtattagt ttgttttggt ctgctgtgat 74160
ggaatgccac agactgtata atttataaac agaaatttat taggacacag ttctggaggc 74220
tggaagtct aagggtccagg tgccagaaag ttcagtgtct ggtaggtca ttctcttcca 74280
agatggtgcc ttgcatgctg cctagatcct cacgtggcag aaggccaaag ggagaaagag 74340
gatgagccca ttacttttt ataatggcat taatctatc atgagggtc aacctaaatg 74400
cttcccatga ggccccactg tcacattggg aattaagttt ccaacacacc aattcttggg 74460
cacacattca aactgtagca gtaggcaaag ccaaatagtg gaaggctttt attctaagct 74520
gtgggactca gctttattct gttaacagtg aagatgcaga aaggcgaga tgcagtggct 74580
catacctgta atcctagcac tttgagaggt tgaagtgagt ggatcacttg agccgggag 74640
ttcaagacca gcctgggcaa catagtgaga tccgtctct atgaaaattt ttaaaaagta 74700
aaagttagct aggtgtggtg tcacgtactg ttcccaacta ctcaggaggc tgaggcaaga 74760
ggatcacttg atcctgggaa gttgaggctg caataagcca tgatcacacc actgcactcc 74820
agcctgagtg acagcaatac cctgtaccac ttccccacca aaaaagatgc aaaaaaggga 74880
gatttatctc cctagttttc taaaggagca aaagcaactt ctgtgggtgg aatagcccta 74940
taagtagaag gggccaaaag ataaaaagg caaccaaga agggaatgag gggatgagaa 75000
gaagacaggg actcatatct gcacattagt ctagggaata aagaagaaa gacacaacat 75060
caatctaaaa agaaataaat gaaaacaata aaaggcagt ttgaaggcag tcgagcactg 75120

acttctatag tatgggttaa agcccagctt cacctctaac tttctctgaa atctttaaatt 75180
gttctaacct ttcttatttt gtcatctat aaaatggaga ttttatctca gtttggttg 75240
agattaaaca agttaatgat tctaactact tagaaccttg ccaaggatat aataactgct 75300
taataacccc tcacaatcat gttctgtgtt ctagagaagt cagggttaagg gtagtaattc 75360
tcttttaaac catgaagtct acctagagtt atgtaggatg tcaccattgg gggatgcttt 75420
gcggtaggca catgaaacct ttctgtacca tttattgtaa catcttgtga gtctctaatt 75480
atttcaaaat aaaagtttta ttttggtttg tttttcttaa agccttttca aaccacaatg 75540
gcaattcatt caagcatgtt ctcagcactt gctgcatgcc atacctgag ctcttctgt 75600
tctaacatta gtaaacctca ttttccaata ttcagaaagc aattaaccaa atgacagtgt 75660
gactttccat ttttactatt gctctttagc agaaacatag caataaaaat aaactattgg 75720
aatgaaatt atgagtatcc tagattaatc cccaaagttg cagtttcttt tgttggttcag 75780
ctggttgatt ttctaagagc ctgaaccag ccaagtgc atctatcttc tgagcagagt 75840
tctgcccagt tctctgtctg actagaattg cacagaaacc acaaactctc cataaagcct 75900
aatgggtgtt ctcaacagag gcatgattgg aattctgagc tggataattt aatactaggg 75960
tacagcccat gggtagtaca gaaaagagtc aatacttgcc ttgtctttgt acaaccagga 76020
tgtatgtctg catgacaata atgaagactc aatttcagcc ccattcacc cttccacaac 76080
agcaggggca gggcaccccta gagcataaat aatctgaaca accctgtgtt gggggcctgc 76140
tgaagatgct taacatccct ggcccatcc actaattgtc tgccatcaat cacaacttga 76200
cagttacaag taaattgcct gtgttctccc catatgcctg cctttctcat tgtgtaaaaa 76260
cagctctacc tccatctgct agtcactacc aatcactctt gccaagatgg tgactcctct 76320
tcttgccctg tagtccctag acacaaagag tccaaagtgc cctgggtggca gccatagctt 76380
ttagttcaat ggtaccttg ctatgttccc tggcagaagt acacttcctt tggtgaccag 76440
gacctccaag ttgaagagca cagaggtgca gatgaagaag tgcaaagccc gttagtgggt 76500
caatgggagt gatgttaact taggccattt ctctctgtac cccttgattc taagacctat 76560
gtatacttcc tgtaggggac ccaacaccat gaagtgggct ctgatttaac acatatgctg 76620
cactttaaag gatgggaccc taaactttca gagcattacc ttaagtggc acttcatatg 76680
ttccttcagt agttcattct agggctctaa aagtacagct acttctgtct ggtgaagtat 76740
gtaatacaaa cagtaggttc taggatcaag agctatccta cacccttc acaattaagt 76800
gggtcccctg atcacatact atgttggtgg atttccattc ttgtttggta atctaaaagc 76860
ccccagataa cagttagctg tgttgactat gactcagaga tcaggaaagg aaagcttata 76920
cccagaataa gtgcatcatt ggaaggatgg acagctggct gatgcaggac agaggggaac 76980

caatgtagtc aacttgccac aaatcaacca gtcgaatctc ctgaagacct agtaacctat 77040
tgagggctcc ttgttggtct ctcttgtag caagttggac attcaaaatc agcagtagct 77100
agatttacca ggataattgg tagtccattt tattggctaa tgagtagtct tgctctctgc 77160
cactgtggcc actcgattca tgtgcccac atgccagtac ttaatccaaa gacaaagcct 77220
ggataatgtc aactggacaa gttatcttgt ctatttggtt tttcagtgct tcttgtgtac 77280
ttgatgcttt ccaattagtg ttaaacgcat gacacaaaaa tcttcacaca ttgtgccac 77340
tcccatagtc ccaaccacat gtctctgtct cagactcttg tttccaatct ccaagtcct 77400
ttcattcagg tccctgtcca cacagctaga ctattggcta ctgtgcagaa acctatacat 77460
atattcacac cttaggccaa tttttcttcc acacgaatag ataaccatat tcattgcttg 77520
cagcttcatt cattgggaag accttccttc tccactgtct tccaaggcta cccataatt 77580
cagctgtact tcagtgatca tctatctttg ctttacagct acttaaggaa ctaaccatt 77640
gtcagccagg attggccttt tacctccctt tcagctagtc atagagaacc cccatatagc 77700
cataagtgtg agctatgcaa gggcactggg acaaccctgg tgggtgacat ggaggcatct 77760
gtgctacttg ctcatgcagc ctcttcagtc cagtctagtc ccatatgggt ttgatcccag 77820
atgtgccatt tccattttcc aatagaaggc tgtgggacac gtctgagtgt atgacttagt 77880
gcattggcaa taatctagct cataatggtc agttccaaat acatttggtg cccattatc 77940
agctgttcca tcttgaccaa ggctcagcaa cattctagga agggtttctc aaaagggtgta 78000
taactttgta ctgaaaatgg atggcattgc ttcagagccc taggggcctg cattctgatt 78060
ctttcgcttg ggcattccac aactccaaac tgtatctttc occactaaca cattcagcac 78120
catggggtcc accagatcat atgcctcaag tagtagggtt gcttatactg cagcccatac 78180
atgctgcaga gccctttctt gctttcttgc tccaggccca ctcaaagctg tgagacttct 78240
atgtagttca gtatatggga ccaaaaatat tcctagatgt ggaatgtgtt gcctccaaaa 78300
caciaagagg cctaccagg atcttgctct atttattgag attgagaatg taagatacag 78360
caatttaact ttacttgaa aatgcagtgg tgggagtggg attctagcag actttcacta 78420
ctgaaccact aaaaacttta ccgatgtggc atgttctga attttcacag ggtttgtctc 78480
tcaccctcca gagtgtatgt gtccaccac tgccctaat gtattagaca cttcttgttg 78540
actcttaatg acatctttga tgtaaatgga tcaatgtgat attctacaga atgttcagat 78600
ggccatgtc tcttcagacc ttattatgac agatgtttga agaggtaaca tagccatggg 78660
caaagcagaa aacgaatatt attgtctgtt tcatgtgaat acagactctg attttttggg 78720
tgtgtctttg tctggttttg gtatcaggat aatattggcc tcgtagaatg agtttggaag 78780

tattccctcc tctctatatt ttcagaatag tttagagtaag atgatattag ttcttcatta 78840
aatgttttgt ataattcagc agagaagcca tcagttcctg aacttttctt tattggaaga 78900
cattttatta taagttcttg ttatttgta ttggtcatag ttcaatcttg gtacattgta 78960
tgcacttagg aatttgttcc tttcttctgt aatttccaat ttatttaciaa tttatttatt 79020
tgagtcttct ctcttcttta ttagtctggc taaaggtttg tcaattttgc atgacatttc 79080
aaaaatccaa ctttttggtt cattgatatt ttgtattgtt ttcttcatct caatttcatt 79140
tatttctgct ctgatattta ttatttatct tctgctacta attttgagtt tgttatgctc 79200
ttccttttct agttctttaa gatgcatcat taagttgttt atttaaagtt tatcttcttt 79260
ttcaatgaag gtactaatag ccataaactt cctcttagt gattttgctg tatcccatag 79320
attttggtat gttgtgttcc tattatcatt tgtttcaaga aacttttaaa ttttcttctt 79380
aatttcttca ttgaccacc agtcattcag gagcatatgg ttgatttcc atgtatttgt 79440
atagtttcca aaattcctct tgttattgat ttctagtttt attccattat ggtcagagat 79500
gatgcttcat attatattat tttattttgt tttaatgttt tcaggctttt tagtgattta 79560
acatatagtc tactcttgag aatgaccat gtgctgtgga aaataatgtg tattctacag 79620
ctgttggttg gtctatagcg cagattaact ctctgttttc tttaccaatt ttctgtctgg 79680
aagatctgtc caatggtgaa attgtggtgt tgaagtctcc agctattatt gtatggggtc 79740
tatatctctc tttagctcta acaatattta atttatatgt ctgggtgttc cagtgttggg 79800
tgcatatata tttacaattg ttatattctc ttgctgaatt gaccttttt tcaactatata 79860
gtaaccttct ttatctcttc ctatagtttt tgtcttgaat tctattttgc ctgatataag 79920
cataactagc cctgctcttt tgtggtttcc attgacatgg aatatctttt cccatccctt 79980
tattttccgc ctacatatgc cttatgtctt tattggtgaa gtgtgtttct tgtaagcaac 80040
agatcattga atcccttttg ttttttcaat ccgttcagcc acttggtgac ttttgattgg 80100
caagtctcag agtctcatcc aaagtcttca atgtacctgg gtattgctgc tggttattct 80160
gagcccaggg actctttagt tagcaggtga tgaatgttgc caggactatg tccttccctt 80220
caaggcagta gtttcccttc tggcctaggg catgtctaga aatgccatcc aggagctaga 80280
gcttgaaaag gcggcctctc agagctgact agtgccctct cctgttgtga ctgtgctgg 80340
atccaagatg taagacaaaa tccttcttac tccttcttc tccttcttct cctttcetta 80400
agtgaagaa agagacctct tttggagctg tgagaattgc agcctggggg taggggaggg 80460
gtagtgccag aactccctta gccaccaag ctggtatctc agtagtccat gtgcctcccc 80520
agtgtactgg ctctggggcc agttcagaac tagaacttgc ttgaaagttg cactccttgt 80580
ggcctagact gactttcaag tgtatttaga gccctagagc actttagctt gcagtggtaa 80640

ggcttggtggg aactcaagtt ctgaccacca ggattggtga tttccttttg gctagagcta 80700
atataaatgc tgtctccatg ggggtggggca tcagctgagt ttggcccagt tttcttttat 80760
gctataatag gacagcacta agttcaatgc ctcacaattg ctgtgctttt cccccctggc 80820
acctaggaat gctctccaca cctgctgcc attggtgggg gatgggagag ggggtggcgtc 80880
agagatttaa aactgttttt tttttttttc tacttcttca gtgcctcttt cagcaatatg 80940
aagttaaac caagcactat gagagctcac ctgatttttg gtgcttaca agetactttt 81000
tttgtgtgta gatagttgct aaattagtg ccttgctgag ggaggaagag ggaacgatca 81060
gtggaacctt ctattctgca acctgcctg cagcactctt ttgagaaaga ttctggatta 81120
cgtgataaat tagtgatgtc tgtcatggag ggggtgagggg cagtttttag acatatagcc 81180
tatcttgcca tttctgctat aaacatctgc acagcttgta caatctgtaa ctatagaatt 81240
gagtacagat gacctgcca caagaggaaa aatgtcctga tcttggcctt agtgcctctg 81300
taattcaacc aagtatatta attaaatgga tagatcatc acttgctgtt tcccaagtgt 81360
agaggttgaa agtggaagg agaagaaagc aaagaaaact gggaggaggg gcactaagac 81420
tgggcataag gagaacaag aggtccaatt tgcaaaataa tgggtgtagt aaggggaatt 81480
ttgaagtatt ctcaaagat aacgattttg cttttctcaa ttttgcctc ttgattctca 81540
cacagtgcct tcttttagtac agacctacat accccgaaaa cttctgttct gtgacttaa 81600
aagtcaactc tatgattagc atgattcctc ttcagtcatt actgaaatc ttcctaggtg 81660
gttacattta agtggtgaaga agcttcagta atataagtca aagaactcct accttgcttt 81720
gcatttcctc agtgggcctc tgataaagca ggaaggaaaa aaattaaaaa tgaacctccc 81780
caggatttca tttctattgt ggcttagctg caaatttgaa aaataaaaat aaaaataaac 81840
cctgaatgta agcacttgta cgtttttggt agattctttc acaggactgg tcaacatctt 81900
gacagatatt aattggcgat tcaactggaga cttcacggca cctgacctgg tttgccgagt 81960
ggtcgcctat ttgcaggat gtcacacctt ccaaattgta tgagacaaac tgataccaga 82020
gttagctttt agatttatct aaggaaaatc atataaaacc ttgcattcct atgcatcaaa 82080
ctgcccgtt totcctgggt agaagggcaa tgaaaatttg atttttaatg tatttccttt 82140
gctaaaaata gtatctatat aaagaaaaca caaacgagcc acaaatgagc aatgctaagg 82200
acataaagca ttcatatttc aaatgttaaa taaaatgtct gggccaatat taaggatttt 82260
ggcatattgt ggataatctc caaaaagttt gttaattata tttacctttc acttctcaaa 82320
tgaactcaac gctggctact gctctttttt gtaaaccaag gggaatttgt tggaacataa 82380
ccattaacta ttgaattact gcctaaatta tactttttta agtatctatt ggcatttgaa 82440

ttttctcctc tcaaatttcc tattttcctt tttgaatgat tttataactg agagactttt 82500
tagtttttct aagatctatt tttagattaa agatattgat gcttcaaag tcaatctgct 82560
tcattgtaag ccatatttat gtatcatatg tatatttgat ttaaaatatt acatatatac 82620
atatttacat atatacatatc ttgttaatat gcattagatg tttataatat agatttagac 82680
acttcaacta agtctggaag gaatggatta tttcataact ggtattaggt ttgatgtgct 82740
gtttctgagt tcttaagatt attttttatt aatatatttt tattctgatg ctagtacttc 82800
aggtctaaat tattatagct ttatgtatat ataatgcatt tctagaaagc tttatcctgc 82860
tagattccat ttgttagtag actcttttag gatctttgta tatatattca taagtaaaat 82920
tttaattttt ttcattgggt ttattatttt caggatcatg taatgaagtt attccatatt 82980
cagaaaatga atatggaata gttacaatct tcttttctgc tctgaatata acctgaaag 83040
tttggtatgt ctgaaagagc cttttctatt atgtacggct agtattttct caggggaaat 83100
ttttaaaaac atatttcaat ttctttctga ggatatttgc ttccttgtaa ataaatgctt 83160
tcaattgctt attatattag caatcaattg ctgatacaca aattatgaca aaacatgggtg 83220
gcttaaaaca ataaacattt attatctcac agtttctatt agtcaagaat tcaggaacag 83280
cttggttagg cagatctgga taaaggctta tcaaaagggt cggtcagatg ctttgtggta 83340
ctgtagtcat cagaacgctt ggctgggact ggacaagatg gatcactcac ataggtggta 83400
aattagtact ggctttagtg gaatactcag ttctccacat gggcctctct ataggtacat 83460
tttagtgtcc tcacaacata gtagctggct tttccaagaa taacctattc aagtgactga 83520
ggtggaagtg gcaacatatt tatgaactag tctcaaaggc catacaccat cactttcata 83580
tgattatttt ggtgtgacaa atcaacttgg attcagtatg ggaaaagaag aaccacaaaa 83640
ctaggaagta aggattattt tcgacatggt ggaggttggc aaccacacta tcttcatatt 83700
tttccaaagc atttgataac ctatggtttc tccaattgt taactttaaa atgtcttctt 83760
gatttagaaa tattcatgtc agtaaaattt tttctaactg tatctacaat aaatgttaag 83820
actatatata aaatatatct tttgacatag ggaaaacgat agagaaagaa attctgataa 83880
ttctctcttc tataaaggaa gcaagaaaat tggcaaaaaa tatcagaatc aaatttgtaa 83940
aactctggaa attaaccaaa cacttgcaat tactcagga gcatttatcc aggaaaatgg 84000
ctgaatctca gtgagaaaaa tgagctttgc ggtgctttgg cttgcattat tctcagcccc 84060
cagtatccaa ctctgtaaca gccttaaaaa gtaatagctc tcattcatag tgaaaacaaa 84120
ttgtctggca gccactgtag aaagcaaaac agggacagag ccctttcaaa tttccaagaa 84180
attgccatta tttgacatgt ctggcagttc tctggaagac tccacttgca aggtgtctt 84240
tattttactt aaatcagaac tcaccactg tgaaaaacgt ttctcctctg ggcattcgtc 84300

aaaacagtta caggcaattg attaacctttg cttatgacaa aggtattgga taatagtcag 84360
ggaaaacaac agactaacca aagcttaaga acatacgctg agtaatgaga tatccatagg 84420
ggatttgaaa atctctgaca tactcccaaa aatctagaag accaggtata tgcccaggac 84480
tatgctcatc ctaaggaaaa aatagagaag gcactgagct gtaacctgtg aatgatcttg 84540
aggctttgtg caagcaggaa gtgagggcaa agacagaatt gtaagttgcc tggctaagtg 84600
tcaaaatcat gctccaacat tcacacagac ttcctctgca atgaatgggg gaattaatag 84660
tttctagaat gtaagaaaat ctctgttttag gacctatcac caagatcatg gagtaggaga 84720
tattcatccc tccaaagcat tcaagtatag acagctattc acaaactaaa atagcccaga 84780
gaggactcaa gggcccatga aagaatttgc agcaacatcg tgaaggaaaa aatggagaat 84840
aaccacatag aaagaattgc tgggtgtgatt ggcacaccta agaggccatg atttggctag 84900
gaacaaagaa gaaaggtgga agctatcagt attagacaca gaaaggacca acattgtctc 84960
tagtggcctg ctccacagga gacactagca tcttttccca ctgagtaacc aacagtcatt 85020
cctgcaagag aaccccgagag ggggagatgt ggctataoct tctcctatc cccaagaag 85080
cagttgctgt cgagctgctt tgggaatggg gtcaccacct ctccaacca tgtgtaagct 85140
ccagccctgg agccatggcc accctgggaa tgcccacact ccagactgat agtctgacta 85200
cactgggctc acccatgttt ttaacatcac agccagcacc atagtaagct agttaacact 85260
ccaagcccca gtgttaaact ctgcctgcat gtgtccacac tccagacact gacttaacca 85320
ccatagagag ctagcccat ccaggcctca gagccattat aattctgtgc atacctgtgc 85380
tctcattctt gtctccttga ctaatttata cgtatactat taccaacatt acagcagggg 85440
cacctgtgcc ctgggcaaca gctcagccat agaaagctac gccttgcccc aacctgaag 85500
ccactgaaag actgtgttct cagctccac actactttgc atgtgcactc atgcctcaca 85560
tactatacca tgcagcagaa ggggcatctg caccttgagc accaatgtca ttaccattct 85620
aatcccagag ccataatctc tccatgtgag cccatgcac agacctcagt ttcattggcta 85680
ctccatgagt gtcaccatc agacactggt gatactgcca ctaagagagg ccctacaagc 85740
cagatctagt accaagagaa atcttctcag ctccagcttc cctgggtggg ggaaagatca 85800
gaagaacctt agcagccatc accactgaag accctatcaa ttcttgctgc tactgccaat 85860
atccactgct ttggccactg aggatcctta tgatctttat caactctgac ctgagctgaa 85920
aaagctacac agagactaca aatctgcacc atcactgatg ctacccacc caacaaacct 85980
tccctcatag ggcattggtt ttccatagtg aactagtcca taaaatctgg aaggtgattc 86040
tttttgggtt tggtaggtt tatctttccc ttatattgat gtgtgtgcat ttgactgctt 86100

taccaaacgc acacacatca acataaggca aaaataaaca taaaaaaca gacataacat 86160
catcaaaaga acacagtaat ttacagtaa ctggcccaa tgaaatagag atatatgcaa 86220
tgcttgacca aaaaaatcaa aataattttt ttaggaaagc ttagtgaact taaagtacag 86280
agaagcaatt caatgaaatc agtaaacag tacaaccaa acaagaatt aaagagagag 86340
actgaaatta ttttttaatt aaatagaaat tctggagcta aaaaacatta taaatgaaat 86400
aataaatgca ctagacagca tccagcagaa attatcaagt agacaaagaa tttgtgaagt 86460
gaaaaacagg ttatttgaaa atatacaact aaagaaaaag aaagaaaaaa gaatgagaaa 86520
agatgaagaa agcttaagag atttatgagg cagcatcaaa agagaaaatg tgcaatgcat 86580
tagtccattc tcatactgct ataagaatgt attcaagact gggtaattta taagggaaag 86640
agctttagtt gactcacagt tctacagggg tggggaggcc tcaggaaact tataatcatg 86700
gtggaagggg aagcaaaccat gtccttcaca aggtggcagg agagagaaga atgagaactg 86760
agcaaaaggg gaagcccctt ataaaaccat cagatctcat gagaacttac tatcatgaga 86820
ataccatggg ggaaaccacc tccatgattt aattacctcc cactgaattc ctcccatgac 86880
atgtagggat tatggcaact aaaattcaag atgagatttg ggtaggaaca tagccaaact 86940
atattataca agttatagga gttatagaag gagatgagag agacaaaaaa aatatagatt 87000
atttaagaa ataatagcag aatagtttgc aaatctgggg aaagatacaa atatccaggt 87060
gtacaagaag gtcagaagtt tctaatacaga tttaatgaaa ataagactac accaagacat 87120
attataatca aactgtcaaa aatcaaaac agagaaaatc ctgataacag caagagagaa 87180
gaggcaaacc acatataaga gagttataac aaggctagca gaaaatttct cagcaaaac 87240
cttgacagaa aggagagact agaagaacat atgtaaggaa ggaaaaaca actgccaaac 87300
aaaaatatat acccaacaaa gctgtccttt agaatcaag tagagagaaa gatttcccca 87360
gacaaacaaa agctgaaaga gttcatcacc accaaacctt tttccaaaa attctacaca 87420
gaatcttcaa gctaaaggaa aaaaaaaga gcgagctaat tagtgacatg aaaatatata 87480
aaagtgattt tgggctgaga caatgggggt ttctagatat acaatcatgt catctgcaa 87540
cagggacaat ttgacttcct cttttcctaa ttgaataccc ttcatttcct tctcctgcct 87600
aatggccctg gccagaactt ccaacactct gttgaatagg agtggtgaga gagggcatcc 87660
ctgtcttggt ccagttttca aagggaatgc ttccagtttt tgcccattca gtatgatatt 87720
ggctgtgggt ttgtcataga tagctottat tattttgaga tacatcccat caataccgaa 87780
tttattgaga gtttttagca tgaaggggtt ttgaattttg tcaaaggcct tttctgcatc 87840
tattgagata atcatgtggt ttttgtcttt gggtctgttt atatgctgga ttacatttat 87900
tgatttgcat atattgaacc agccttgcat gccagggatg aagcccactt gatcatgggtg 87960

gataagcttt ttgatgtgct gctggattcg gtttgccagt attttattga ggatttttcc 88020
atcaatgttc atcaaggata ttggtctaaa attctctttt ttttgttgtg tctctgccag 88080
gctttgggtat caggatgatg ctggcctcat aaaatgagtt agggaggatt ccctcttttt 88140
ctattgattg gaatagtttc agaaggaatg gtaccagttc ctcagaaaaa atgctcacca 88200
tcactggcca tcagagaaat gcaaatcaaa accacaatga gataccatct cacaccagtt 88260
agaatggcaa tcattaaaaa gtcaggaaac aacagggtgct ggagaggatg tggagaaata 88320
ggaacacttt tacactgttg gcagaactgt aaactagatc aaccattgtg gaagtcagtg 88380
tggcgattcc tcagggatct agaactagaa ataccatttg acccagccat cccattactg 88440
ggtatttacc caaaggacta taaatcatgc tgctataaag acacatgcac acgtatgttt 88500
attgtggcac tattcacaat agcaaagact tggaaaccaac ccaaagtcc aacaatgata 88560
gacgggatta agaaaatgtg gcacatatac accacggaat actatgcagc cataaaaaag 88620
gatgagttca tgtcctttgt agggacacgg atgaaattgg aaatcatcat tctcagtaaa 88680
ctattgcaag gacaaaaagc caaacatcgc atgttatcac acttagatgg gaactgaaca 88740
atgagaacac atggacacag gaaggggaac atcacactct gggcactgtt gtgggggtggg 88800
gggagggggg agggatagca ttaggagata tacctaatac taattgatga gttaatgggt 88860
gcagcacacc agcatggcac atgtatacat atgtaactaa cctgcacatt gtgcacacgt 88920
accctaaac ttaaagtata ataaaaaag aaagaaaata tataaaagta tatatcactc 88980
actgttaaaa gtaagtacat ttacaatact ctaataatgt aatggtggtg tgtaaatcac 89040
tgatatcatt aggataaagg ctaaaagaca aaactattaa aaataataat agctacaata 89100
gaggtatacc tcagagatat tgtgggttca attccagaat gcctaaataa agcaaatttc 89160
acaataaagc taatcacaat atttttgttt ccagtagcat ataaaagtta tgtttacatt 89220
ttactgtagc cttttaaatg tgtgataata tgtctaaaaa gtatacatgt taatttaaaa 89280
atacttaatg gctaaaataa aaagattaac agtcatctga gccttcatca aatattggcc 89340
tttttactgt tggaaagtct tgctttgatg ttgatgactg ctgacttatc aggggtggtg 89400
ttgccaagg tttggtggc tgtggcaatt tcttaaaatc agacaacaat aaagtttgtt 89460
gcatcaattg acttttcctt acatataaaa aattatctgt agtatgcaat actgtttgat 89520
agcattttac ccagagtaga acttctttca aaattggagt caattctctc aaatcttgcc 89580
actgttttat caaatatgtt tatgtaotag tttaaattat ttgttatcat ttcagcaatg 89640
ttcacagcat tatcaccaag agtagattcc atttcaagga atcactttct ttgctcatcc 89700
ataataagca attcctcatc cattctagtt ttattaagaa agtgaagcaa ttcagtcaca 89760

tgtacacact ccacttctaa ttctatttct cttgctatct ccaccacttt cgcagtagct 89820
ttctccagtg aaatcttgaa cccatcaaag tcattcataa gagttggaat caaccttcta 89880
cactcctgta aatgttgata cattcacctt ctcccataaa tcatgaatgt tcttaatgac 89940
atctaagtag tgagtccttc tggaagggtt tcaatttccc ttgtcctgat ccagcagagg 90000
aattactatc tatggcagcc atggccttac aaaatgtatt tottaaataa taacactgga 90060
aagtcaaaat tactccttga tctatgggat gcaaaatgga tgttggtgta gcagaaatga 90120
agacaagttt aatctttttg ttcatcgtca tcaaagctct tgggtaacaa tgtgcactgt 90180
caatgagcag tgatatttgg aaaggaatct ttttttctga gcagtagata tcaacaggga 90240
gcttaaaata ttcagtaaac catgctgtga acagatgtgc tgccctttag gctttgttgt 90300
tttatttata gagcacaggc acagtagatt agtctaattg ttaagggcc taggatcttt 90360
agaatgatca atgagcattg gcttcaactc aaagtcacca cctgcattag ctgttaaaaa 90420
gaatatcatg atgtcctttg aagctttgaa gccaggcatt gacttctcct ctctagctat 90480
agaagcgcta gatggcatct tcttctagta ttttactgtt tgtttccact gaaaaactat 90540
cgtttactgt agccaccttc atcagttatc taacctcaat attctggata acttaatgca 90600
acttctgtat tatcacttga tgctttacct tgcatgttta tattacagag acaacttcca 90660
tccttaaacc tcatgaaaca aactatgcta gcttcaaact tttctgctgc agctttcaca 90720
tctatctcag ccttcataga ggtgaagaga gtttaagaac ttgttctgga ttaggctttg 90780
gcttaatggt gcaactagtt tgatcatata tccagaccac taaaactttc tccatatcag 90840
cagtaagctt gtttcatttt cttgtaattt gtgtgttcac tgaagtagtc attttacttt 90900
tcttcaagaa agtttccctt acattcaca cttggctaac tggcccaaaa ctaagctttt 90960
ggctctatctt gggttctgac atgctttttc tccctaagct aagtcatttc tagcttttga 91020
tttaaagtga gacatgtgtg actcttccta tcaactgaac acctaaagac aatttcaggg 91080
ctatttattg gcttaatttc aatattactg tgtctcagag aataggaaca gctgaggaga 91140
gggagacaga tggaggaaca accagttggt ggagctgtca gtatgcacac aacatttate 91200
agttaagttt gtcaccttat atggatgtga tttacagcac cacaaaacaa ttacaatagt 91260
atcatcaaac atcacttgat cacagatagc cttaatggac gtaataataa caataatgaa 91320
aaagctcaaa ataaccaaaa tatgacacag acatgaaata aacacatagt gctggaaaaa 91380
tgacagactt gctcaacaca gggttgccac aaaccttcaa tttgtaaaaa aaatgcacta 91440
tctgtgaaga acaataaagt aaatcatgac aaaatgaggt atattttagt tgtgttaagg 91500
gatatactat ataaaaaatt taaattgtga catcaaaaac taaaatgggg gtggtagagt 91560
aaaagtgtag agtttttgtg tgtaatcaga gttaagttat cattttaaaa tcgctcgtta 91620

taagacattc ctttgtaagt ctcaaggtaa ccacaaagca aaacctgtag tagatacatg 91680
aaagataaaa agtaaggaat caaagcatac cattagagaa aagcatcaac cacagaagaa 91740
gacaggaaaa atagaagaaa ggaaaaaaaa cacacaaagc agaaagagtg gaagacagaa 91800
aaaaaatata tgatagccag aatataatta ttaaaatggc aatttagggc acatataact 91860
gaaagtaaaa atggaaaaaa gatattccat gcaagtggaa accaaaatag atgaggggtt 91920
gctacattta catcacatag aatagacttt aagtcaataa cagtaaaaca gggcaaataa 91980
ggtcattgca taataataaa cggattatth catcaggaag aaataacagt tataaatata 92040
tatggaccct acactggatc acctaaatat ataaagcaaa tattaataga cctgaagaga 92100
gaaatagact gcaataaaaat cataggtctt caaaatccca cttttaataa acagattatc 92160
tagacagata atcagtaaag aacgattgga tttaaactac acttttagacc aaatggaccct 92220
aatagacaca tatagaacat ttcacctagt ggcagcagaa tatataatct tttgaagcac 92280
ccaaggaaca ttctccagga taatcataga gtaagccaca aaacaagtca caacaaaatt 92340
aagattgaaa tatataaatt attttttcaa aaacaatgac atgaaactag taatcaataa 92400
tataaggaat ttcaaaaatg ttacaaacgt gaaaattaaa caagagactc ctcaacaatc 92460
aataggtcaa agaagaaatt aaaagcaaaa atttaaaata tcttaagata aatgaaaatg 92520
aaaacacatc ataacaaagc ttatgggatg caacaaaagc aggtctaaga ggggagttha 92580
taccaataaa tacctacatc aaacaaaaga tttaaaataa ataacctaat attacaactt 92640
aagtaactag aagaagaaga agagtcaggg tttccctttc ataaaaaggg agaaataata 92700
aaaattagag aagaaagaaa acaaatggag aatagaaaaa caatagaaaa gatcaatgaa 92760
gagttagtth tttgaaaaga taaaattgac aaacctttag ctagaataac taggaaaaag 92820
aagaaaactc aaataaaatt agaaatgaaa gctgagacat tacaactgat accatagaaa 92880
caaaaagcat cataacaagc gactatgaaa aactaaatgt cagccaggcg tagtggttca 92940
tgcttgtaat ccagcactt cgggaggctg aggcaggcag atcacctgag gtcaggagtt 93000
tgacaccagc ctgaccaaca tgcagaaacc cgtctctac taaagataca aaaattagcc 93060
aggtgtggtg gcgggcccct gtaatcccaa ctactcagga ggctgaggca agagaattgc 93120
ttgaaccag gaggcagaag ttgcagtgcg caaaaaaaaa aaaaaaaaaa aaaaatagat 93180
aacctagagg aaatggataa gttcctagat acacaaaaac taccaagact ttgagtattg 93240
aaggaatgga aaatctgaac acaccaatac taaggagatt taatcagtaa taaaatgttt 93300
ccattccaag aaaagaccag gatcagatgg cttcactgca aaattctacc aaacacttag 93360
agaagaacta atagcaatct ttctcaaagt cttcccccaa aaaagtgaaa aaagtgttha 93420

cttccaaact catgttaca agtcagcatt actctgataa caaagtcaga caagatacta 93480
tgagaaaatt acagggcaat aaccctgatg aacatagggt taaaaatcct caacaaaata 93540
ctagcaaaca tatttcaata gcacacttga aatattattc accatgatta agtgtaattt 93600
attcctggaa tgcaacaata gtttaacata cacaaattaa taaatgtgac atgccacatt 93660
aacaaaatga aagataaaag gcatatgatc atcttaatag atgcagaaaa acatcgtgac 93720
aaaattcaat atctttttgt agtaaaaatg ctcaacaaat taggtacaga agagaagtgc 93780
ttcaacaaat aaagaccaca tatgacaagt ccactgctaa catcacagtc aacagtggta 93840
aactaaaagc ttttcttcta agattaggaa caagataaga atgctcacta ttgcaacttc 93900
tattcaacat agtactggaa gttctagcca aagaagtaag gcaagaaaaa aagaaaagaa 93960
aaaagaaagg cattcacatc agaaaggaca aacttaaatt ttctgtttga agatgacctg 94020
atcctacata tatataaaat atataattta tatataatat gtattatgtt ataaatatat 94080
gttatatgtt ctatgttata tataatatat gttctatgtt atatataata tatgttatat 94140
gttatatgtt atatataata tatgttatat gttatatata atatatgtta tatgttatat 94200
atatgttata gggtatatgt tatatatgtt atagggtata tggtatgtta tatatatgtt 94260
atatgttatg ttatatatgt tatatgttat attatatatc atatatgtta tatgttatat 94320
tatatatcat atatgttata tggtatatta tatatcatat atgttatatg ttatattata 94380
tatgttatat gttatattat atatgttata tggtatatgt tatattatat atgttatatg 94440
ttatatatta tattatatat gttatatgtt atatattata ttatatatgt tatatgttat 94500
atatattata tataatatgt tatataatat aatatatatg ttatatatta tatataatat 94560
gttatatatt atatatgtta tatattatat tatatatgtt atatgttata tattatatta 94620
tatatgttat atgttatata ttatattata tatgttatat gttatatatt atatgttata 94680
agttatatat aatatagata atatataaaa tatataatat atatacacia aacacaaaac 94740
accacaaaaa actgttaaaa ctaacaaatg aatttactat agttgcagga tacaaaatca 94800
acatgcaaag tcagtaatgt ttctatacac taacaatgaa ctatccaaaa aaaaatagaa 94860
tctatttaca ataactccaa aaagaataaa atacgtagaa atgaagttaa tcaagtaggt 94920
gatagactta tacaaaaaaa ctataatata ttgatgaaag aaattaaact aggcaaaaat 94980
tggaaggca tcctgttttc ataaatatac agacttaata ctttaaagtg totataaatc 95040
aaagctatct acagattcaa tgcattccct atcaaaatcc caacagcatg ttttacagaa 95100
attaaaaaaaa atcaaaaatt caaattaaat cacaaaagtc ccagaattgc cagtgttaac 95160
ttgagaaaga acaaagctca aggcattcaca cttcctgctt acaaaaatat attatgaagc 95220
tacagtaatc gaaacagtgt tgtactggca tgaaggcaga catatagaca gatgaaacag 95280

aataaagaag ccagaaataa atccatgtat gtatggtcca ctgatcttca acagtgatgc 95340
caaaaataaa cagtgggtgcc aaaaataaac aatggggaaa ggacagtctc ctcaacaagg 95400
tggttgaaaa ctggacatct acatgagaag agtgcaattt aacccttaac ttatacaata 95460
cagaaaaaaaa tcaactcaaa acagattaaa gacttaaatg taagacctta aactataaaa 95520
ttcctagata aaaacagggg aaatgttcct tgacatttct cttggcagtg attttttttg 95580
ataagacacc aaaaactcaa gaaacaaaac cacaataga caagtgggac aacattaaac 95640
taaataactc tgggtgcaata aagaaaacaa tgaatcaaat gaaaaggatga cctatggaaa 95700
gggagtaa atttacaagc catacatatg ataaagggtt aatatccaaa atatattaga 95760
aacatctaca actcaattta aaaaatagta ccaggcatgg tggctcacgc atatagtcct 95820
agtacttttg gaggttgagg ctggtgaatg gattgagcgc aggtattcaa gaccgccag 95880
ggaaacatgg caaaccccta tctctacaaa aacacacaca cacacacata tatatatatg 95940
tacatatata cacacacaca tatatatgta catatatata tatgtgtgtg tgtatatata 96000
tacgtatata tgtgtgtgta tatatgtgta tatgtgtgtg tgtgtgtatg tgtatatata 96060
tatgtatata tatatttata tatatgtata tatatatatg tatatatata cgtatatata 96120
tatatttata tatatatata tatattagcc aggcattggtg ggtgtgtgtc tgtagtccca 96180
gctactcagg aggctgaggt agtaggattg cttgaacccg ggggtgtcaag gctgcagtga 96240
gccatgacag caccattgca cttcagcctg ggcaacagag agagaccctg tctcaaaaca 96300
aacaacaaaa aacataggct gtgtgcagtg gctcatgctt ataataccag caccttggga 96360
agccaaagca ggaggatttg agaccaggag tttgagacca gcttgggaaa catagcaaga 96420
ccctgtctct acaaaaagta aaaataaaaa taaataacca ggcatatttg catgagttag 96480
tagtcctagc tacatgggag gctgaagtga gaggattgct tgagcccagg agttcaagac 96540
cagcctgggc aacatagtga gacccccatc taggaaaaca aaaagcacag gcaacaaaag 96600
caaaagtata caaatgggat tgcaggaaac taaaaagctt ctgcacagca aaagaactat 96660
caacagagta aagatacagt ttacagaatg ggagactatt ggcaaactat gcatctgatg 96720
aggagttaat atccagaata tatgagaaac tcaaacaaact caatagcaaa aaacaaaaaa 96780
taaaccat ccaaaaaatg ggcaaatgac ctgaatagac atttctcaaa aaatatatat 96840
acaaaagaag acatacaaat gggcaacatg tatacgaaaa tatgcttaac atccctagtc 96900
atcagggaaa tacaatatcca aaccacagta agataccacc tcagtctagt tagaatggct 96960
actacagaaa gacaaatgat aacaagtgtt ggcaaggatg tggagaaagg aaactcttac 97020
atattgttgg tgggaatgta aattagaaca gtcactataa aaacagtatg gaagttcctt 97080

gaaaaattaa aaattgaact gccatatgat ccagcaatcc cactaatggg tataaaccca 97140
aaagtaatga aatcagtatg tcaaagagac atctgctgtc ctgtgtttat tacagcacta 97200
ttcataatag ccaagatatg gcaacaatct atgtccatca atggatgaat gaatttataa 97260
atcttggtat atacataata tgcaatacta ttcattcata aaaaagaaag aaattgtttc 97320
atctgtgaca acatagatga acctggagaa cattgtgtca agtgaaataa gctacacaca 97380
gaaagaaaac tactgcatga tttcactgat atgtggaatc ttaaaaagtt gatctcatag 97440
aagctgagag tagaatggtg gtttagcagaa gctggtgagg gtaagtgaag gagggatagg 97500
gagaacactg gtcaatgagt acaatgttaa agttaaatag aggaacgggt tctagtgttc 97560
tattgcacat caggggtgaat ataaaaaca atagtgtatt atatatctca aaataacttg 97620
aagaaaggat tttaaatggt ctcaccataa agaatgata aatgtttaag atgatggata 97680
ttctaattac cctgatttga gcattacaca atatatacac atatcaaac atcatatcgt 97740
accacataaa tatgtacgat tattctggct caattgaaaa taaaataaaa ttttagaaac 97800
gttaagaaag gaatattatg aacaatctta tctcacagat tttataacct agttaagatg 97860
gaccaattcc ttgaaagata cagtctatta taactcatac agggagaaat aggcaatcta 97920
aatactgtta attgtattaa aggaattaat tattgaataa acttccaaaa cagaaagcat 97980
caagcccaga tgggttcact ggagtattcc accaaacatt taggaaagaa aactacacca 98040
agtctctaca gttgcattga aaatatagaa gcagaggcaa tattttctaa ctcatgttat 98100
aaggtcagca ttacactaat accaaaagca gacaaaaata ttgcaaggaa aaaaacacat 98160
caatatatct gatgaatata gatgcaatat ttctcaagaa atattagcaa gtcaatccaa 98220
caaatttata aaaataattc tacacttgag ataagcctgg gcaacatagc aagacactgt 98280
acctacaata aatacaaaaa ttagccaagc atgggtggtg atacctgtac tcctagctac 98340
ttgggaggtt gaggcaggag gattatttga gccagggggt tagaggctac agtgagctat 98400
gattgtgcca ctgcactcca gcctgggtga cagagtgaga ccctatctct taaaaaaga 98460
aaaaagaatt atacaccaca aacaagtaag gcttatccca aaaaagtaag actggttcaa 98520
cactggatat cagttaatgt aaccattat atcaataggc taaagaagaa aagtcacaag 98580
attataacaa cagattttta aaaagtgttt aacaaattac aacacctatt atgatttcaa 98640
aaaaaattca gtaaaactagg aatagaaagt aatcttcctc aacttgaaaa ataacatcta 98700
caaacagcc ttcactgtac atcttattta atagtgaag actatgagct ttcttgctaa 98760
gattaggaag aaaacaagga tttccctctc acattccttt tcatcattgt acagtaagtc 98820
ctagctagtg caatgagaca aaaaaggaaa taaaagacac actaattgga ggagagaaat 98880
aaaatgttgc aggtgacatg attgtctatg cagacaatct aaaataatta accaaaaaaa 98940

ctctcccaga actaataggc aattaaagca agattacaga atatgttaat atacaaaagt 99000
tcattgcttt cctatataac cgcaataaac aagtataatt tgatattaaa aaacaatatt 99060
atctgcatta gcacacctga aatgaaata attaggtata aatttaataa aatatgtgca 99120
agatctatac aaggaaaact acgaaacttg aatgaaagta atcaaagaaa aactaaataa 99180
agagttattc catgtatata taggaagatt caatattgtc aagggtgcag atcttctcaa 99240
cttactctat gtagattcaa cacagtccca gtcagtaacc cagcaagtta ttttgtggat 99300
atcaacaaac tgattctaaa ctttatgtga aagacaaaac atgcagaatg gtcagcataa 99360
tattgaaggt gaagaacaaa tttgaaggac tgatacttcc caaaatcaat atttactata 99420
aagctatagt aattaagaca gtgtgctatt agtgaaagaa caaacaggtc aatggaacag 99480
aatagcaagc ccagaaatag tctggcaaaa atatagtcaa ccaatattta caaaaaagca 99540
aaatacaatg gagaaaaggt cataatttta acaaatgggtg ctgaaacaag aagacatcca 99600
cacgcctaata aataaatcta aacacagatc ttatatccat gacaaaaacg aacataaaaag 99660
aaaacctaga tgaccttgag tatgggtgatg actttttaga tacaacacca aaggcatgat 99720
ccatgaaaaa aagaactgat aagctggcat ttattaaaaat taaaattaaa atttctgctc 99780
tgcaaaaggc aatgagaag caaaccacag actaaagaga atatatttgc aaaaggcata 99840
cctgataaaa gattgttatt caaaatatgc agaagagctc ttaaagcaca ataggtaaat 99900
gaacaatctg atttttaaat ggtcaaaaaa acagacacct caccaaaaaa ttacacagaa 99960
ggcaaataag taaatgaaaa gatgctcata tacttatgaa tatctttggtt attagggat 100020
tgcaaattaa aacaatgata tagcactgca tatttattag aataatcaaa atctaaaaca 100080
ttaacaacac aaaatgctat tgagaatgtg aagcaacagg aacgttcact cattgctagt 100140
gggaatacaa aatagcacag ccacttcagc gacagtttag tggtttattc caaaattaca 100200
ttcttaccat gcattccagc aattgtgctc ctgggtatctt acccaaataa gttgaaaaca 100260
tatattccca caaacacctg catacaaacg tttatatcat gtttattcct attttccaaa 100320
cttggaagca accaagatgt cctttggtag gtaaatggag aaataagctg tgggtacatcc 100380
atacaatgaa atattattca gtgctaaaaa gaaatgagct gtcaagctta tttaaagggtt 100440
acgtgtctca tgcaaagaca tggaagaaat ttatgcatat tagtaagcaa aagtgcagct 100500
ctgagaagct acatactgta gattccaact ttatgacatt ctggaaaagg caaaactctg 100560
gagaaagtga aaacattagt ggttgccagg gcttggggag agggaggcag agcacagagg 100620
gggaacaaag ggcagtagaa ctactctgta ctatactata acagtggaca catgtctata 100680
cttttgtcaa aatacaatgt ataacaccaa gagtgaatcc taatgtaaac tatgaacttt 100740

gggtgataat gatgtgtcaa tgtaggttca tccactgtaa caaatatacc actcttgcac 100800
ggaagcgtga tggtagtga ggctctgtgt atgggggaca gggggtttat gggaaatctg 100860
taacttctgc tcaattatgt tattaatcta aaactgctct aataataaaa tctactttta 100920
aacatgtaat agaattccca agaactgtgg gacaattaca caaagtgtag cacacacata 100980
taacagaaat accaaaaaaa gaaaagagca gaagaagtat ttgaagaaat aaaagatgaa 101040
aaattttcta aattaatggc aaacaccaat ccagagatcc aggaagctca gagaacacca 101100
aagcaaaata aatgtcagaa aaaatgtacc tctaggcata tcatattcaa actgcagaaa 101160
accaaagaca aagataaaat cctgatagaa gttgaggggg aaacttacag ataaataaac 101220
aataataaga atttcctctg acttatcaga aatcagccaa gtaagaagat agtataataa 101280
aatattttaa gtgggaaaag aatacaaatc cccctgacct agaattctct atcaagtga 101340
attattcttc aaaagtgaag aagaaataaa gactttctca gacatacaaa aacaggaaat 101400
ttctcaaaag taaacttttc ttacaaaaaa tgctaaaaga agttcttcag ataaagggga 101460
aaatgataaa ataataagca caaatcaata taaaggaaga gcaattgaga aggaataaat 101520
gaaggcaaat taaagctact ttcttactc ttaattgacc tatagatagt tgctcgacac 101580
aaaaatagca ataatgcatt ggtaatcatt attaccactt attaggataa gtgaactgaa 101640
tgacaatatt aaaagagatg cagcccaggc acgggtggcc catgcctgta atcccagaac 101700
tttgggaggc cgaggcaggt ggatcacttg aggtcaggag cttgtggcca gcctgaccaa 101760
catggtgaaa ccccgctgtc actaaaaata caaaaatagg ccagatgtgg tggcacgcac 101820
ctgtaatccc agctactcgg gaggtgaag caggagaatc tcttgaaccc ggaaggcgaa 101880
ggttgcagt atccaagatt gtaccactac actccagcct gggtagacaga gtgagactct 101940
gtctaagaaa aaggggaggg ggggtgaatg caaagagaaa tacaatttaa ctgaacagct 102000
atcagttaaa gtgaactgaa ctattaatta aatggatcta attggcattt gtagaatact 102060
tcatccagta atagcagaat atatcttctt atcaaaacta tatgcgattt ccatcaagag 102120
aaacttcata tggaccataa aacacatttt aacaatgtaa aaaaatagaa attacacaaa 102180
gtatactctc agaacacaca gaatagaact aaaaatcagt aacagaagga tagctgaaaa 102240
ttttcaaact gtttggagat tgaacaacac acttctaagt aataagtga ttgaagaaga 102300
tatctcaaga aatattttga aataaatgaa aatgtaacat catagtttgt aggatgcagt 102360
aaaagcagtt ctttgagtga aatttataac actgaatgga tatattagaa aataaaaaaa 102420
tgtgatatta ataaactaat cttaggcctt aggaaactag agaaagaaaa gaaatgtagg 102480
cctaaagaaa gcagaataaa agaaatgata aaaattggag cataaattat taaattgaag 102540
tagaatttaa attttctctg cagagtaagt atgtgaggta atgaatatgt taattagctt 102600

tacttagcca ttccacatgt atacataaac acgttatata gcataaacgt atattatttt 102660
gtttttcaat taaaaattta attaaaatca gaaaaaatta ataattaaat ttaaaacagg 102720
aattcaatac tgaaaaatca atgaaaccaa aagctgtttc tttaaaaaga taataaaatt 102780
gataaatttc tagccaggct aatcaagaaa agaagaggat tcaacattct taaagaaaac 102840
aattttcaac ccagaatttc atatccagcc aaactaagca tcataagtga aggagaaata 102900
aaatatttta cagacaagca aatgctgaga gacttttgtc accaccaggc ctgccctaaa 102960
agagctcctg aaggaagcac taaacatgca aaggaacaac cggtagcagc cactgcaaaa 103020
ccatgacaaa atgtaaagac catcgagagt aggaagaaac tgcataact aacgagcaaa 103080
ataaccagct aacatcataa tgacaggatc aaatttacac ataacaatat taaccttaaa 103140
tgtaaatggg ctaaagtctc caattaaaag acacacactg gcaaattgga taaagagtca 103200
ggacccatca gtgtgctgta ttcaggagac tcattctcatg tgcagagata cacataggct 103260
caaaataaag ggatggagga agatctacca agaaaatgga aagcaaaaaa aagcaggggt 103320
tgcaatccta gtgtctgata aaatagactt taaaccaaca aagatcaaaa gagtcaaaga 103380
aggccactgc ataatggtaa agggatcaat tcaacaagaa gagccaacta tcctaaatat 103440
atatgcaccc aacacaggag caccagatt cataaagcaa gtccttagag acctacagag 103500
agacttagac tccaacacaa taataatggg agactttaac accccactgt caatattaga 103560
cagatcaacg agacagaagg ttaacaagga tattcaggac ttgaactcag ctcagacca 103620
agcagactta atagacatct acagaactct tcaccccaa acaacagaat atacattctt 103680
ttcagcaccg cattgcactt attccaaaat tgactacata attggaagta aagcactcct 103740
cagcaagtgt aaaagaacag caatcacaa aaactgtctc tcagaccaca gtgctatcaa 103800
attagacctc aggattaaga aactcactca aaaccgcaca actacatgga aactgaataa 103860
cctgctcctg aatgactact gggtagatag tgaaatgaag gcagaaataa agatgttctt 103920
tgaaaccaat aagaacaaag aactatgta ccagattctc tgggacacat ttaaagtagt 103980
gtgtagaggg aaatttatag cactaaatgc ccacaagaga aagcaggaga gatctaaaat 104040
tgacacccta acatcacaat taaaagaact agagaagcaa gagcaacaa attcaaaagc 104100
tagcagaagg caagaaataa ctaagatcag agcagaactg aaggagatag agacacaaaa 104160
accttcgaa aagtcaatga atccaggagc tggttttctg aaaagatcaa caaaattgat 104220
agactgctag caagactaat aaagaagaaa agaggggaaga atcaaataa tgcaataaaa 104280
aataataatg tggatatcac taccaatccc gcagaaatac aaactaccat cagagaatac 104340
tataaacacc tccatgcaaa taaactagaa aatctagaag aaatggataa attcctggac 104400

acatacaccc tcccaagact aaaccaggaa gaagttgaat ctctgaatag accattaaca 104460
ggctccaaaa ttgaggcaat aattaatagg ctaccaacca aaaaaagtcc aggaccagac 104520
caattcacia ctgaattcta ccaaaggtac aaagaggagt tggtagcatt ccttctgaaa 104580
ctattccaat caatagaaaa agagggaatc ctccctaact cattttataa ggccagcatc 104640
atcctgatac caaagcctgg cagagacaaa aaaaaaaaaa gagaatttta gaccaatatc 104700
cttgatgaac atcgatgtga aaatcctcaa taaagtactg gcaaactgaa tcctgcagca 104760
catcaaaaag cttatccacc atgatcaagt cggcttcac cccgggatgg aaggctggtt 104820
caacatatgc aaatcaataa atgtaatcca gtatataaac agaaccaaaag aaaaaacca 104880
catgattatc tcaatagatg cagaaaaggc ctttgacaaa attcaacaac cttcatgct 104940
aaaaacactc aataaattag gtattgatgg gacatatctc aaaataataa gagctatcta 105000
tgacaaaacc acagccaata tcatattgaa tgggcaaaaa ctggaagcat tccctttgaa 105060
aactggcaca agacagggat gccctctctc accactccta ttcaacatag tgttggaagt 105120
tctggccagg gcaattaggc aggagaagga aataaagagt attcaattag gaaaagagga 105180
agtcaattg tccctgtttg cagatgacat gattgtatat ctagcccat catctcagcc 105240
ccaaatctcc ttaagctgat aggcaacttc agcaaagtct caggatacaa aatcaatgta 105300
caaaaatcac aagcattctt atacaccaat aacagacaat cagagagcca aatcatgagg 105360
gaacacccat tcacaattgc ttcaaagaga ataaaatacc taggaatcca acttacaagt 105420
gacgtgaagg acctcttcaa ggagaactac aaaccactgc tcaatgaaat aaaagaggat 105480
acaaagaaat ggaagaacat tccatgctca tgggtaggaa gaatcaatat cgtgaaaatg 105540
cccatactgc ccaaggtatt ttatagattc aatgccatcc ccatcaagct acaaatgact 105600
ttcttcacag aattggaaaa aactacttta aacttcatat ggaacccaaa aagagctcac 105660
atcgogaagt caatcctaag ccaaaagaac aaagctggag gcatcacact acctgacttc 105720
aaactatact acaaggctac aataacccaa acagcatggt actggtacca aaacagagat 105780
atagatcaat ggaacagaac agagccctca gaaataacgc cgcatatcta caactatctg 105840
atctttgaca aacctgagaa aaacaagcaa tggggaaagg attccctatt taataaacgg 105900
tgctgggaaa actggatagc catatgtaga aagctgaaac tggatccctt ccttacacct 105960
catacaaaaa tcaattcaag atggattaaa gactaaattg ttagacctaa aaccataaaa 106020
acctagaag aaaacctaag cattaccatt caggacatag gcatgggcaa ggacttcatg 106080
tctaaaacac caaaagcaat ggcaacaaaa gccataattg acaaatggga tctaattaaa 106140
ctaaagagct tctgcacagc aaaggaaact accatcagag tgaacaggca acctacaaaa 106200
tgggagacaa tttttgcaac ctactcatct gacaaagggc taatatccag aatctacaat 106260

gaactcaaac aaattttacaa gaaaaaaaca aacaacccca tcaaaaagtg ggcgaaggac 106320
aagaacagac acttctcaaa agaagtcatt tatgcaggca aaaaacacat gaaaaaatgc 106380
tcaccaccac tggccatcag agaaatgcaa atcaaaacca caatgagcta ccatctcaca 106440
ccagttagaa tggcaatcat taaaaagtca ggaaacaaca ggtgctggag aggatgtgga 106500
gaaataggaa cactttttaca ctgttgggtg gactgtaaac tagttcaacc cttgtggaag 106560
tcagtgtggc gattcctcag ggatctagaa ctagaaatac catttgaccc agccatccca 106620
ttactgggta ttacccaaa ggactataaa tcatgctgct ataaagacac atgcacacgt 106680
atgtttattg cggcactatt cacaatagca aagacttggg accaacccea atgtccaaca 106740
gtgatagact ggattaagaa aatgtggcac atatacaca tggaatacta tgcaaccata 106800
aaaaaggatg agttcatgtc ctttgtaggg acatggatga aattggaaat catcattctc 106860
agtaaaactat cgcaaggaca aaaaatcaaa caccgcaggc tcgcactcat aggtgggaat 106920
tggacaatga gaacacatgg acacaggaag gggaacatca cactctgggg actgttgtgg 106980
ggtgggagga ggggggaggg atagcattag gagatatacc taatgctaaa tgacaagtta 107040
gtgggtgcac cacaccagca tggcacacgt atacatatgt aactaacctg cacattgtgt 107100
acatgtaccc taaaacttaa agtataatag taataaaata aaatttaaaa aaagatagca 107160
tgagtagaaa aaaaaaaaaa aagaggggaag ccacaagtta ctaatatcag aaatgaaaga 107220
ggggtcatca ctactgattc tgtggacatt aaaagaatga taaataaata ccataaaca 107280
ctctatgccc acaagtttga tagcttagat gaaatgaacc tttccttga aaagctcaat 107340
ctgccaaaac tcacataaag tgaaatagat aatttgaata ggcctatacc tactaaagaa 107400
actgaattaa taattaataa tcttccacaa aagaaaacac caggcccaga tgggttcatt 107460
agtgaattct gtcaatcatt taaaaaaaaa atggttaaca ttatctgcag tcttctctag 107520
aaaacagatg aaaaggggaa cattcctaatt tcatatactt gctgacttgt tccatgaggc 107580
cagaaatgcc ctaattctaa aaccaatgat attaaaagaa aggaaaacta cagaacaata 107640
tcctcatga acataaatgc aaaaattctc aataaatatt agcaacttaa atttttgttt 107700
aaaggataaa tcatgtatag aaagaattat agtccgtggc caagtgagat ttattccagg 107760
catgaaaatc tggttaaata tttaaaaatc aattaatata atctatcaca caaacaagat 107820
aaagaaggaa aattatgtga tttttccatc aatggacaca ggaaaagcat ttaacaaaat 107880
tcaacatcca ttataataa aaactcagca aactggaaac aaagagaaac tgctcaact 107940
tcattaacag catctaaaaa aaccatacag ttaacctcat atttaatggc aagaaattag 108000
gtgcttttgt cctaagactg ggaacaaggc aaagatgtcc totctcgcca ctcccattca 108060

acactgaact agatgtctta gataatttaa caggacaaga aaagtaaata aaaagtttat 108120
acattggaaa agatgaaata aaaatgtctt tgtaataaaa tttgtaattt tctatattaa 108180
aatctgaaag aatagacaag aaaaataaaa acaaacaaaa aagaaacttc cagacctagt 108240
aggcaaaaaa atggcaacac tgaagaatag aaggtaata tacaaaagtc agttgttttt 108300
ttccagcaat taccaagttg agattgaaat taaaaacaca ataccattta tgtgaacacc 108360
aaaaaatgaa acttaggtgt aaatctaaca aaatatatac aagatctatt gaagacgact 108420
acaaaactga tggaaaaaaa agcaaatatc taaataaatg aagagatgtt ccatgttcat 108480
gaataggaag acttaatat gttgaatgt ctattctgcc caacgtgatc tataaattca 108540
atacaattcc aatcaaagtc ccaacgagat attttatcaa tataaacaaa ctctaaaatc 108600
atatggaaa gcaagagaca cagaataacc aatgcaatat taaagaaaa gagcaagttt 108660
gaacaaccaa cactacctga ttttaagact tactataaag cttcataatc aagacagcat 108720
ggtattggta aaagaatcaa caagtagatt gatgaaacag aacagagagc ccagaaatag 108780
acatacaaaa agatagtcaa ctagtctttg acaaatgagc aaaggcattg agagatagcc 108840
ttttcaagaa tggctgtctt agttgaaaga gtggttgaat atccacatgc aaaaaaaaa 108900
aaaaaaaaa aacttaatgt agaccttatg cctttcacaa aaaccaactc agtttagatc 108960
atagatataa cataatgctc aatactatta tataaacga ctagaagata acataggaga 109020
agatttaagg gacctactct ttgtctatga gtttttagat acaacaccag aagcataatc 109080
catgaaaaat gattgaaatg ttggacatta tgaaaattta caactttgca aaggacactg 109140
tttaagagaa tgaaaagaca gccctcagaa tgggagaaaa tatctgcaga atatatattt 109200
cataaaggat ttgtatccaa atatgtacaa agaactctta aaactcaaca ataagaaaac 109260
aactcatttc aaaaacggcc aaaagatctg aataaacctt tcaccaagaa gatggcaagt 109320
aacagatag aaagaagctc aaaatcatat gccattaggg aattgcaaat taaaacaaca 109380
gtgagacaac actacacatc tattagataa tcactaaaat tcaaaaaaaaa ctgaaaatac 109440
caatggctga tgaggatgaa tattaacagg aactttcatt cattgctgct gggaattcaa 109500
aatagtacag tcactttgga agataattta gcagtttctt acaaaactaa gcatggtctt 109560
accatacaac ccagcaattg taaggctatg tatttactca gttgagttga caacttatat 109620
ccacagaaga atgtttattg tagctttatt cataatcacc aaaaactgga agcaacaag 109680
aggttcctca aaagctgaat ggataagctg cggtacatcc attaaatggg atattattca 109740
atgatataac aaaatgggct atccagccat gaaaagacat ggaggaatct taaatgcca 109800
ttgctaagt aaagaagcca gtctgaaaa gctacatact aaatgattca actatattac 109860
attctcaaaa aggcaaaact gtagaagagc aagaaaccag tgttgccaga gatttggagt 109920

aggaaagaga tgaatagata aaagcagtgg attttttagga cagtgaaact cttctgatat 109980
tataatggtg ggtacataac attatgcaat ggtcaaaacc catagaactg tccaacacag 110040
agtgaacctt aatataaaca atggacttta gttaatgata atatatccat attggttcat 110100
cagttttaac aaatgaacta caagatgtta atgacaagga aaagtatata gatgggaggg 110160
agtcataggg aaaccctctg tgttgtttgc aattttttgt aaatataaaa ttgttctgaa 110220
aaataaagtc tattttctta aagtcagtc agggatcctc ctttatatga atacaatctg 110280
gcattttgct aagtaaaatt tctctgtttc cagcatttaa gtctcacttc tctgtgcaag 110340
attttcttta gttttatatt tcctaggtca tttctgtcag aattaagggg tggggtagt 110400
agaattcaga catatgtcct caagctgcca tctgttctcc aaatgggatc ccatatgttg 110460
ttgatgtgta gagttttcat tgttgacatt tatgaattga ctataattgt actttgattt 110520
ccttcttgac tcgagttaat tatgaggggtg tttaaatttc caagtgggat tttttaaaaa 110580
aacatgttac taatttctgc ttcattgctc attgttttaa atatgtgatt tgcgtaagtc 110640
ttacttttta gagctattga gaggggtttt tttcttggtg gtttagtata tcgccaattt 110700
ttgaaatgct tagatatatt atagcaatgt gtactattct tatgttatga agtctatttt 110760
tttctatctg tatatttttg tttattcttc tgtaaaactt tgaacggttt cattatatat 110820
ttagcattat gttacttggt acatataact ttgtgaatat cctattgcca tcatgtgttg 110880
tatcttttat caatataaat aatcctttat cacatgaaat gctgtttgct ttgaatttaa 110940
tttgtctaag agcaatgttg taatgaatgc cttctttaat ttttactttc ttgatatacc 111000
tttggccatc cctctatttc cacacttaga tatctaagtg atgtggttta aaatttttct 111060
cttttaaaaca gcatagaatt agatttttagg gaagttttta aataagttat tttcttaatg 111120
aaagacotta agtcattcac atgtattgog acaaatagtg ttttggttg gtgtaatttc 111180
agttccttat tttagtcttt ctatattaag ggttttttgt ttgtttgttg ctatttactt 111240
gttttttctt tttatttgat tatcaattgt ggtaggttg aaaaaagctc cccgccccta 111300
ccccccacca ccaccaccac cacaaaatcc gtatcctaata ccttgggaacc tgtgatgtta 111360
ccttatataa gaaaaggga cttcacaat gtaatcaaat taaagatctt aaaactggga 111420
gattattctg gtttatccca gtgattccca catgtaatca taacagggtg gtagaaggga 111480
gtgtgacaca cagacagaag aggaaaaggc agagcttggg ccgatatggt cacacgcaa 111540
ggactgctgg cagtcacccg aagccagaag aggcaagggg cagattctcc cctagagtct 111600
ctggaggact catgggcttg ccaaactttt atttttgccc agtagaacta attttggact 111660
tctggtctcc agagctgtgc gagaataaat ttatgttgtt ttaagccata aagttttag 111720

tagtttgta cagccaccat aggaaactaa ttaaaatttc tccattttta ttttactcta 111780
tcatagctta tctttaactt ttagctttaa tatactctta tcaagggttaa aaaacaaaag 111840
agagacccaa aatacataac aaaattgaac atgatgtctt tctttgtact ctgtcccttc 111900
actcaactcc ctttaagttt tcttgaaatg tccatagtc atatttataa ttattttgta 111960
agcatttcat gtgtttctaa cttcattgct taccatagtt gttattatc tatgctcttc 112020
cctcttcttg aattcttaat ttttaattcaa ttctcaatta gctgaagtat gccaggaaga 112080
ctataaaata ttatattttc ttccttctta tatactctgag aaaactgttc tgtgccttgg 112140
cacagaaatc atatttaca agttcttgaa tcacaatttt ttttcaaac totgaacatg 112200
ttgttatagt gttttctggc attacttttt gtgggaaaga ggtatgataa ttttaagatgc 112260
atttctttta ggtaactttt aaattttttt accctgcctt aaatgcctga agatgggttaa 112320
tctctgaaat taaatactct tttgattgta tctcaaagtc agcatctttt tatcaacatt 112380
gactgctctt ctggacactc ttatttatgg ctcttcctcc tgctccagat tctgaactgt 112440
gcctgtcacc tcctcaaga cagccacaga gcccgaagat ccttataacc tcaccatgat 112500
gttccactcc acagggttct gccattgtgt gctttctttt tcatccttct ctgatgccac 112560
ctcctcctct gtttgttgcc agccttaata attttaaatt catggatcac aggtagaac 112620
cccagttttg tctaagtaga cgatcgattt ttttttctt aaagctagtc ttccattttt 112680
ctctttctct atttgtcacc ttttcccaag agataatttt ttcttttagt tttcagagga 112740
gccttcaatt tcttttttta aacctttttt ctaactttta ttttaagttg aggggtacat 112800
gtgcaggttt cttatacagg taaacttgtg tcatgggggt ttgtgtgaca gattatttaa 112860
tcaccagggt attttttttt tttttttttt ttgagacgga gtctcgtct gtgcgccagg 112920
ctggagtgca gtggcgcaat ctgggtcac tgcaagctct gcctccggg ttcacgccat 112980
tctcctgcct cagcctcccg agtagctggg actacaggcg cccgccacta cgcccggtta 113040
attttttgta ttttagtag agacgggggt tcaccgtttt agccgggatg gtctggatct 113100
cctgacctg tgatccacc gcctggcct cccaaagtgc tgggattaag cttcatatc 113160
attagttatt tttcctgac ctaaccctcc tcccacccta caccctcta taggccccag 113220
tgtgtttgt tcccctctat gtatctgtgt gttctcatca gttagcttcc acttataagt 113280
gagaatatac agtattttgt tttctgttct tgcattagtt tgctgaagat aatgccctcc 113340
atccatgtcc atccatgtcc ctgcaaagca cattatctca ttctttctta cgagccttca 113400
atctctaate attcctcaga acctcctcc tatgggtggc ttcacaacat tagtgctgag 113460
atgcatttgt agaaataaca aatcataaag aaccacaatg tgcttgattg ttatttgtct 113520
gagccagaga aagaacaaaa aggcacgaat tcagaacaaa gataactgcc aagaagaaag 113580

tgaagtaa at gctotccatt agacttcaat ctacaagtta ctcttctact ccctctggat 113640
ttgaacctaa tgttacaaag tggagatata gccagactct tgatagaaat gaggctccaa 113700
gttacctgaa aaatgtccac cttgatcatc ccttcaaata tatttcattt aatcacattc 113760
taaataatgct cttcacataa gagaaattaa cttotttaa at cccaatttcc ctttctgcta 113820
attatgcaaa aaaagattgt acacagataa ttaaaagggtg tgggttaaagt taataaaaga 113880
gacagtggga aggaagagga gaccagatc ctgagcctaa tgaagtggat gaaaaagggg 113940
agctctctga tgggtgtcagt gctgatgatg ataatgatt caacaccgca cataggctcc 114000
ctaaaaccgg aacccttcta tcatctcttg actccactg ataaaaagca aatgaacttt 114060
tgtttcta at tattctggaa tgtcaattgt ggcagaaaaa attagtttag ctttaaaagt 114120
caattccaaa caaatgtacc atgcaaatgt caaggactga tcttgaaata tggagtctcc 114180
aagaggccaa ggaataagt gagcaagaaa gaactgaagc attgtcactc atcatggcca 114240
gcagtacaaa ttctagcacc aggaaccaa ctctgcttcc tgggtgtggag atgaaaaggg 114300
agagtccct ggtccatcca tcacaggaca tgggacaggg gtatggctct ctgttcagct 114360
tcctcaagct caaactcctt ataggagggg gagcatgcag aagggtagggt gcaggagcca 114420
gggcaagtgc attgaactcc ggccccacag tagtgtctgg ggtgggtgcc tgtgactcct 114480
gaagccccag tgttatagt ctgttttaac tctgccatct gcagatggct taagtggtaa 114540
ccagctcagt gccctcttgg tggccatgtc ctctccata ggccaggaag agtcagggtca 114600
cacacagact tgaaggatga atgcaggagt tttattgagt ggtggagggtg gttctcagt 114660
ggatggatgg ggaactggaa gggggatgga gtgggaagat gatcttcccc tggagtcttg 114720
ctgtccagg ccattctcct gtctgattgt cccagctga actcttctta gtgttcagac 114780
cctccttctc ttctctcctt ctctgccatg ctgttctgct attcgtctgc tcatctcctc 114840
ctggagcctg ggatttgggg tttatatggg tacaggatag gggcatgtag ctggccaaa 114900
ggcaactttt tgggcacaaa aacaggaata cctgttctca ttaagacca ctggtttcca 114960
ggcttgaggg tgggtcttct gctgggaaac tgcctcttct caccagtat ttccttctct 115020
cttgtccatg taagagagtt atttcccttc ctaatccaac actcaccct tcttgggttc 115080
acctctcaga tttctaaata acacaagggt ctattcttct cattatgtgc cttccctcac 115140
ctctctgaag tcataagctt cctctcattt ctgtctttag gttgtgctgc tctacgctc 115200
tacctacgtc ctgggtgtcc tcagcataga cagataccat gccatctct accccatgaa 115260
gttcttcaa ggaggtgagc tggctttacc aggtgtctct tcataaagaa ctgtccttga 115320
gtgagggtag gatcatttct actaatattt tactctta ataatgtgatc ttcctcctct 115380

ttccttctct tttcttctctc ttcattctcc tcttccacct ctttctcac ttcattgggtc 115440
tgattgtcta aaaagatgta caaaagaaaa cttagatgcc tggaaaaatg ccgatgggtc 115500
ccatggacag cccagagcca ctaaggaagg tgggtggctg acccactttg tgaccattca 115560
gcttgatgct gatgcttcca tctttttcct ctagagcagt gttgacatgc atacctgtcc 115620
aaagaaatga actctccac actaacttct cttttccag aaacaaacac cttgttctcc 115680
ttatccaaaa ccatagaaat aattggaaag atgtttttaa ttgatcatcg tattgggaaa 115740
aaaaaataag gaaactctct ttgatccaga aatggtagg gatccctaag aaaaggaagc 115800
agatgggtca gaggaaggc tgagaccaca cagcccacaa aaccctagg agatgctgaa 115860
gtgggatggg aggccttagg agcacttct gaacaggctg catcaatgtg gagtgtagga 115920
gaggaccga ggcaactgac aaggtgattt ctgaagtaca acacagagga ttaatcacag 115980
caagtttccc tggatttgaa gaggcagaga ggggccctgg ggagcagtaa tgccttagag 116040
gaagggtagg tacatattca caagactagc tgcctactgc tctgcccaca atcagagacc 116100
actctctcc atacaccctg aaagcaaata caaagccatt gtttagcaac ctgttctccc 116160
ttcctccaga taacttcgaa aactgataaa gaacctaa aaaaatcttg accacaaaaa 116220
tgagtgaat caaatattat ttaaaaagtg acaaatcca aactgtagaa aacagataat 116280
acctttaaaa tctcattcca cacaatttgg atctagggaa gcaaatttaa tagaagtagc 116340
agaattttag gataaatata attaacgtca tcaaggatct acaagaagtg gagtgaata 116400
gggtaatttc atttactatt gacaggggta taagtaaaca catcttattg aaatataaaa 116460
tgtgcagccc tttgatatga cttcttggtat tctgctctgg agaaacacc acacgtgtgc 116520
atgtatacat atgtgatgta tattgtggca ctgcctacaa tagcaggaat tcatcgctaa 116580
ccttggtgtg tctcagtgga ggaatggcta actagtctat ggaagaagtt tcagtacctt 116640
ctgtttattg attgcatact aggtgtcatt catgattcta actgattcat acatattaaa 116700
tattttaata ttcataatat ctctgtgagg tgggctctat tattatctgc attttacagg 116760
taaggcaatc gttgttccaa gagtgaagt aatttgccca aggtcacagc tagtaagtgc 116820
aggaacagga tttgaaccca cacagtctgg gtctagctcc agcgccctca aatgctatat 116880
agcctgcctc tcaatgtgga agtggtagag ccatgggcag taatatgcat caacatgaac 116940
atatccatat tgtatgctga gtgacaaaa caatatgcac aaaaatccat gtgaacgata 117000
acattaaagt caaaaactca aaacaatgct agatattgtg tacagaaaga cagagaggag 117060
cagaaagaga gaacacacaa aagaaaactc caggaaaaat aaaaactat ttgcaaaaac 117120
acacaaaact tgacatttga cctcacttct agttatggga gtttagaccag tgggtgtgacc 117180
accacaggaa cttatatgca aggtggtaat ttttacctga agaacatatc aatgtgttgt 117240

gtgcaattaa aatgaattta ttttgatcat gcttggtatg cctactgtac agcataagac 117300
atatgagtaa aaatactaaa ttgtatactt aagaatttgc tggggatctg ttccaagatg 117360
gccaaatagg aacagctctg gtctgcagct cccagcatga tcgatgcaga agacaggtga 117420
tttctgcatt tccaactgag gtgcctggtt catctcactg ggactgggtg gacagtgggt 117480
gcagcccatg gagggtgagc caaagcaggg cagggcatcg cctcacctgg ggagcacaaa 117540
gggttggggg atgtcccttt cctagccaag ggaagctggg gcacactgta cctgggacac 117600
tcctgccaaa tactgcactt ttccaagga tttagaaacg agcagacaag gagattctct 117660
cccatgcctg gctcaggagg tctggagcct tgctcactgc tagcacagca gtctgaaatt 117720
gaactgtgag gcagcagcct ggctagggga ggagcgtcca ccattgctga ggcatgagta 117780
ggtaacaaa gtggcccggc agctcaaact ggggtggagc cactgcagct cagcaaggcc 117840
tactgcctct ataggctcca cctctgtggg caggacatag ctgaacaaaa ggagcagac 117900
agcttctgca gacttaaacg tccctgttcg acagctctga agagcacagt ggttcttcca 117960
gcatggcggt tgagctctga gaatggacag actgcctcct caagtgggtc cctgacacct 118020
gtgtatgcta actgggagac atctcccagt aggggccaac agacacctca tataggtggg 118080
tgccctctg ggacaaagct tccagaggaa ggatcaggca gcaatgtttg ctgttctgca 118140
atatttgctg ttcacagcct ctgctagtga taccaggga aacaggtctt ggagtggatc 118200
tccagtaaac tccaacagac ctgcagctga gggacctgac tgttacaacg gaaactaaca 118260
aacagaaagg aatagcacca acatcaacaa aaaggacatc tacacaaaaa ccccatctgt 118320
aggtcaccaa tatcaaagac caaaggtaga taaaaccaca aagatgggga gaaaccagag 118380
cggaaaagct gaaaattcta aaaatcagag cgctcttctt cctccaaagg actgcaactc 118440
cccgccagca atggaacaaa gctggataga gaatgacttt gacgagttga cagaagtagg 118500
cttcagaagg tcggtataaa caaactactc ctccgagcta aaggagcatg ttcaaaccga 118560
ttgcaaggaa gctaaaaacc ttgaaaaatg gttagacgaa tggctaaata gaataaacag 118620
tgtagagaag accttaaatg acctgaagga gctgaaaacc atggcacaag aacttcgtga 118680
ggcatgcaca agcttcaata gccaatcaa tcaagtggaa gaaaggtgt cagtgattga 118740
agatcaaatt aatgaaataa agcgagaaga catgggttaga gaaaaagat taaaaagaaa 118800
caaacaaagc ctccaagaaa tatggtacta tgtgaaaaga ccaaatctac atttgattag 118860
tgtacctgaa actgatgggg cgaacagaac caagttggaa cacactcttc atgatattat 118920
ccagaacttg cccaaaatag caaggcaggc caacactcaa attcaggaaa tacagagaac 118980
accacaaaga tactcctcga gaatagcaac cccaagacat gtagtcagat tcaccaaggt 119040

taaaatgaag gaaaaaatgt taacggcagc cagagagaaa ggtcagggtta cccacaaagg 119100
gaagcccatc agactaacag cagatctctt gccagaaacc ctacaagcca gaagacagtg 119160
ggggccaata ttcaacattc ttaaagagaa gaattttcaa tccagaattc catatccagc 119220
caaactaagt ttcataagtg aaggagaaat aaaatccttt acagaaaaac aaatgctgag 119280
agattttgtc accaccaggc ctgccttaca agagctcctg aaggaagcac taaatatgga 119340
aaggaacaac cagtaccagc cactacaaaa acatgccaaa ttgtaaagac cattgatgct 119400
aggaagaaac tgcatacaact aacgggcaaaa ataaccagct aacatcataa tgataggatc 119460
aaattcacat ataacaatat taaccttaaa tgtaaattggg ctaaatgctc caattaaaag 119520
acacacactg gcaaattgga taaagagtca ggacccatca gtgtgctgta ttcaggagac 119580
tcatctcatg tgcagagata cacataggct caaaataaag ggatggagga agatctacca 119640
agaaaatgga aagcaaaaaa aagcaggggt tgcaatccta gtgtctgata aaatagactt 119700
taaaccaaca aagatcaaaa gagtcaaaga aggccactgc ataattggtta agggatcaat 119760
tcaacaagaa gagccaacta tcctaaatat atatgcacc aacacaggag caccagatt 119820
cataaagcaa gtccttagag acctacagag agacttagac tccaacacaa taataatggg 119880
agactttaac accccactgt caatattaga cagatcaacg agacagaagg ttaacaagga 119940
tattcaggac ttgaactcag ctgagacca agcagactta atagacatct acagaactct 120000
tcaccccaaa acaacagaat atacattctt ttcagcaccg cattgcactt attccaaaat 120060
tgactacata attggaagta aagcactcct cagcaagtgt aaaagaacag caatcacaa 120120
aaactgtctc tcagaccaca gtgctatcaa attagacctc aggattaaga aactcactca 120180
aaaccgcaca actacatgga aactgaataa cctgctcctg aatgactact gggtacatag 120240
tgaaatgaag gcagaaataa agatgttctt tgaaaccaat aagaacaaag acactatgta 120300
ccagattctc tgggacacat ttaaagtagt gtgtagaggg aaatttatag cactaaatgc 120360
ccacaagaga aagcaggaga gatctaaaat tgacacccta acatcacaa taaaagaact 120420
agagaagcaa gagcaaaaa attcaaaagc tagcagaagg caagaaataa ctaagatcag 120480
agcagaactg aaggagatag agacacaaaa acccttcgaa aagtcaatga atccaggagc 120540
tggttttctg aaaagatcaa caaaattgat agactgctag caagactaat aaagaagaaa 120600
agagggaaga atcaaataga tgcaataaaa aataataatg tggatatcac taccaatccc 120660
gcagaaatac aaactaccat cagagaatac tataaacacc tccatgcaaa taaactagaa 120720
aatctagaag aaatggataa attcctggac acatacacc tccaagact aaaccaggaa 120780
gaagtgaat ctctgaatag accattaaca ggctccaaaa ttgaggcaat aattaatagg 120840
ctaccaacca aaaaaagtcc aggaccagac caattcacia ctgaattcta ccaaaggtag 120900

aaagaggagt tggtagcatt cttctgaaa ctattccaat caatagaaaa agagggaatc 120960
ctccctaact cattttataa ggccagcatc atcctgatac caaagcctgg cagagacaaa 121020
acaaaaaaaa gagaatttta gaccaatata cttgatgaac atcgatgtga aaatcctcaa 121080
taaagtactg gcaaactgaa tcctgcagca catcaaaaag cttatccacc atgatcaagt 121140
cggcttcata cccgggatgg aaggctgggt caacatatgc aaattgataa acgtaatcca 121200
tcacataaac agaaccaatg aaaaaaatca catgattata tcaatagata caaaaaaggc 121260
cttcaacaaa atccaacgcc cttcatgcta aaaactgtca ataaactagg tattgattga 121320
acatatctca aaataataag agctatttat gacaaactca catccaatat catactgaat 121380
gggcaaaaac tggaagcatt ccctttgaaa accggcacaa gacaaggatg ccctctctca 121440
ccactcctat tcaacatagt gttggaagtt ctggccaggg caatcaggca aaaggaagca 121500
ataaagcgta ttcaattagg aaaagaagaa gtcaaatgtt ccctatttgt agaccacatg 121560
attgtatatt tagaaaacc cattgtctca gcccaaatc tccttaagct gataagcaac 121620
ttcaccaaag tctcaggata caaatcaat gcgcaaaaat cacaagcatt cctatacatg 121680
aataagagac aaacagagag ccaaatcatg agtgaattcc cattcacaat tgctacaaa 121740
agaataaaat acctaagaat ccaacttaca agggatgtga aggacctttt caaggagaac 121800
tacaaccac tgcgcaagga aataaaagag gacacaaaca aatggaagaa cattccatgc 121860
tcatggatag gaagaatcca tatagtgaaa atggccatac tgcccaaggt attttataga 121920
ttcagtgcc tccccatcaa gctaccaatg actttcttca cagaattgga aaaaactact 121980
ttaaagttca tatggagcca taaaagagcc tgcattgcc agacaatcct aagcaaaaag 122040
aacaagctg gagacatcaa gctacctgac ttcaaactgt agtacaaggc cacagtaacc 122100
aaaacagcaa ggtacttgta caaaacaga gatatagacc aatggaacag aacagagccc 122160
tcagacataa caccacactt ctacaacat ctgatctttg aaaaacctga caaaaacaag 122220
aatggggaa acgattccct atttaataaa tgggtgctggg aaacctggct agccatatgg 122280
agaaagctga cactggatcc cttccttaca cttatacga aaattaactc aagatggatt 122340
aaagacttaa atgttaaacc taaaaccata aaaacctag aagaaaacct aggcaatacc 122400
attcaggaca taggcatggg caaagacttc atgactaaaa caccaaaagc aatggcaaca 122460
aaagccaaaa ttgacaaatg ggatctaatt aaactaaaga gcttctgcac agcaaaagaa 122520
actaccatca gagtgaatgg gcagcctaca gaatgggaga aaatttttgc aatctacca 122580
tctgaaaagg gctaatatcc agaacttaca aagaactcaa acaaatctac aagagaaaaa 122640
acaaccccat caaaaagtgg gcaaaggata tgaacagaca cttctgaaaa gaagacatct 122700

atgcagccaa cagacacatg aaaaaatgct catcatcact gatcatcaga gaaatgcaaa 122760
tcaaaaccac aatgaaaaac catctcatgc cagttagaat agtgaccatt aaaaagtcag 122820
gaaacaacag gtgctggaga ggatgtggag aaataggaac actttttacac tgttgggtggg 122880
agtgtacatt agttcaacca ttgtggaaca cagtgtggcg attcctcaag gatctagaac 122940
tagaaatact atttgacca gcaatcttat tactggtata taccocaaagg attataaatc 123000
atgctactat aaagacacat gcacacgtat gtttattgct attgcagcac tattcacaat 123060
agcaaagact tggaaccaac ccaaagtcc atcaatgata gactggatta agaaaatatg 123120
gcacatatac accatggaat actatgcagc cataaaaaag gatgagtttg tgtcctttgc 123180
agggacatgg atgaagctag aaaccatcat tctcagcaaa ctgtcacaaa gacagaaaac 123240
caaacactgc atgttctcac tcataggtgg gaactgaaca acgaaaacac ttggacgcag 123300
gggtggggaac atcagacacc ggggcctgtc agggagttag gggctggggg agggatagcg 123360
ttagcagaaa tagctaagt aaatgacaag ttgatgggtg cagcaaacca acatggcaca 123420
tgtataccta tgtatccaac ctgcacattg ttcacatgta caatgataat ttaaaaaaaaa 123480
gaatttgcta acaagataaa ccttatgtta agtgctcttg ccacaaaagg aaaaaactaa 123540
acaaaacaat actaaagggg gacaggaaga agcctttgga ggtgatggat acatttatgg 123600
cattgcttgt aatgatggtt acaggaatgt atacatatct ccagacacat taagtgggat 123660
gcattggata tgtatagctt tttatatgcc aattatacat caataaagtg gtttagaagt 123720
taattttaaa aatagaaaat atacttgttt cctgggtttc ttctcggtca tttggtgtgt 123780
ctgttttgtg aagggttgag acctgtacca agttttctgg gactgtaaa gccctagtta 123840
ttatgatagg ggaggcagga aattgatgag gcaggggaga taggagagac atttcgtgac 123900
ataggatgc aaaaatgtgt caatgaagca tcaaattctg taaaatattt gtagagattt 123960
attctgagtg aaatatcagt gacaatggac aaagccctca ggaggtcatg agaatatgtg 124020
cccaagggtg tcaggggtgca gcttggtttt atacatttta gggcgacatg aaccttcaat 124080
caaacacatt taagaaatac atagggttg tccgaaatgg cagggttggg tagtggggga 124140
tggggtgggc gaagcttcca gattatagtt agatttaaat tttttctggt tgacaattgg 124200
ttgagtttat ctaaagacct gggatcaaca aaaaggaatg tctgagttaa gacaaagaat 124260
tgtggagacc aaacttctta ttgacaggg aaacttttag gtttcctctt caggagagact 124320
atgttgtaaa atgtttctca tcagacttaa agtctgtgtt gatgttaacg ctggagaggt 124380
ataatgagac acacctgaac ccgcactgcc catcatggct tggaacagcc tctcaggtta 124440
aatcttaaaa gagccttggc tgaggaggaa atccattcgg aagggtgggg ggccttataa 124500
ctttattttt gggttacatt catggatttg tgatacactt tttttttttt tttttttttt 124560

ttttagacag agtctcactc tgtcgcccg gctagagcgc agtggcgcg tctcagctca 124620
ctgcaacttc tgcctcccag gttcaagtga ttcttggtcc tcagcctccc gagtagctgg 124680
gattacaggt gtgcatcacc acgcctaatt tttgtatttt agtagagatg gggtttcacc 124740
gtgttgggca ggctggtctc aaactcctgg gctcaagcaa tccaccttgg ccttccaaag 124800
tgctgggatt ataggcatga gacactgcac ctggcccggtg atacacttat ttatccatct 124860
atttgtttat gcctagccga ctttagaaag ggttgaatgt ggtggacaaa aggacatcga 124920
cgccaaagta ttttaagggtg aagaaatata agtgaaggaa aaataagtat aggaaattgc 124980
aggaggggaag tcccaggaaa cttgttagag gtacagactt gagaccata gccagagggt 125040
gacatgatga attaccatgtt tcaccatatt cactagggtta aaaaattagc tgtttgctcg 125100
tggaataat gtctagtctt gggagacact gtgtaatatc atcaagagtg tcctaaatta 125160
tcagaaaggc agcaattaat attatagtgc ttttctaaag atgtaaaaca acttattccc 125220
tcactacaaa attaagagag gctcaacata aggtgcttgg aaaatcctca tcatcagtag 125280
cttctgggaa agatgaaagc ttagaaacac attagccagc tgccttcctc tgcccaggcc 125340
caacgaaagt cactgaagaa atctaaaaac agggggagac ctgcatctgg aatgagagcc 125400
aaggaaagaa gccatccaaa taccagaaac ttctaagaat tcccagcaca cgaggagca 125460
gatgaaacac ggaagccaag tgtataatgt ataattcagc ccccgaggaga actccctggc 125520
cttgtagagg atgaactgat actcacaagt ttatagggtg agggcagccg gtgagaacgc 125580
gacacgagag gttccaggaa gccaacgcg ggagtcattt tagtaccgca gcttgtcaga 125640
atttaagggc agctgagagg cagggttcgg agtttatgcg gctcagctga gagaaagccg 125700
tagaaagaag agattccccg agagtggcag ggctctctgg gagagtcatg gagcaggag 125760
aaaccaggag ggagtcaaat gtagctgtgc aggcacaagg caatggccgc ttaactgagg 125820
ggaggccata ttttatttat tttttactgc tattttactga agcagtattt ttcttagcca 125880
gcaactccag aaaagttgac aattttaaag gtcttcacca aaaatgaata tgagaataat 125940
tagaaaaata tcaaaataat aaagagaaat atgcagtgtt atattaaaat gtaatataaa 126000
actttgataa ttaaaacaat gctggaacca gaataacaca tataaatcat acaaaatttc 126060
aaatcagttg gcaaaaaatg aattagtata cgggtgtgaa atagctgtct ggcctttttt 126120
cagaatatta gatcaccacc caccatagct aataacaaag taaatttttag ttggattgaa 126180
aaaattaaca cttcaaacat aaaaggagta aaagaatatt tgggtaacac cttcataatt 126240
tcagtgaagg aaaggccttt ccaggaacca tatcaaaagc agaagcaaga aggaaaatat 126300
agttttttac tacataaaaa tttaaattat ttatgtaaat aataaatatc cataaacaaa 126360

cttaaaaagc aaaattataa atggaaaata tttttatgta acaggcaaag aaaaatatcc 126420
tttttataaa gaagtctggc aaagaaataa aaacattcta attgagtaaa tgggcaaagg 126480
gcatgaacag actaacaaga aacaaagaaa cataaatgag accaggcgct gtggcccacg 126540
ccggtaatcc cagcacttcg ggaggctgag gtgggtggat cacttgaggt cacaagtttg 126600
agaccagcct ggccaacata gtgaaacct gtctctacta aaaaaatata aaaattagct 126660
gggcatgggtg gcatgcacct gtaatccgc tactcaagag gctgaggcac aagaatcgct 126720
tgaacctggg aagcggaggt tgcggtgagt ggagatggca ccactgcact ctagccctct 126780
agcattggca acagagcaag acttcaccag aaaaaaaaa aaaaaaaca caccataaat 126840
gaatagaaaa tgtatagaaa tattcatttt gactactaat taaagaattg caaaattgaa 126900
catttgata caaaatttta ccaaccacat aacaagatta aaaggaagaa ttatccaact 126960
attaggagag tataaggaaa ttggtacttt tgttgttaca gtgagagcca taaatagtag 127020
gcacttttga gggaaatgta acaatatttg tcaaaattta taaatcttct gtttctaaga 127080
gtttatccta aggaataaca agaactgaga aaaatagatg aatataaatt atttataata 127140
aaaaatggaa acaacttag ataataatcg gctattgata aaattggagt acagtataac 127200
caccggaaaa aatgttatth atcagtgctt ttccaattgc acttcccaca gtgaccccca 127260
gagggtcat tttgctggc tgctccaggg cagatctggg agctgctggc tgtctcagtt 127320
tcaaccaag atacctgagc tttcaacact ttcctatact gggttcgaga taaaattata 127380
tttgaagaaa ggtacttact acctgaaaa aaatgccact tgtacaagtt catatttctt 127440
aacaggaaaa gatggtcata acaggctttt cagaggaaaa aaattagggtt acaaggtagg 127500
atctgtatca tgatctcatt tctaaatgtg taatctatgt attatatatg catgcgttat 127560
caaaatttaa gcttgacgtt tctggttggg gccactatgg atttttttcc ttttcgctta 127620
tttttatthg ctagtgtttc tataatcaat gtgtgtaact tatacgattt tataaaaagg 127680
taaagagtt acatggaaaa gaggaaggct aagcggcaga gaagtacaga aatggacaag 127740
agcacatgga agagggagtc tgactttccg ggggtgagaa acatgagaca gtctacctta 127800
tatgcaaacc tgcatttcat ctgttcgac ctggatcctt cagccctgag aagtcactct 127860
tcatggcaaa tgggctcagc tgcctggaag gccatcagtc tccctgctgt ggcagctgcg 127920
tcacagccta agagagtggg gctttcacag tgtagcccca gcaccagtaa gagaaaaaca 127980
gtcccaagcc tataacctgc tgggcctthg gggaaacatc ttcggagacc ttggtagaat 128040
tgattctaaa gtatgattct atactctctt cctgtccat ccctctcaa agtgaatgca 128100
ctgttcatcc atcaagtgga aactttgtgg ggcccaggag tcacacgttg cccttaagag 128160
tgccctggaa cctatagaag aagatgacat tagaattcta accattgaaa gaacaagagc 128220

acagggggca gctactaatg tcacccagga ccatagaaac ccatcaggct tcacaccccc 128280
 accatcaaga agaaccacct ccagtcctat tacctgccct cttcatccca ccctgcttct 128340
 ttcttaatct gaactccac tgcttctctc cacgcctggt ttttctctgt tccttttaggg 128400
 accagctccc ctatgccatc aaactcctaa tgggaggctc cctccacctc ctgcctcctg 128460
 aggtggctga ttcccttaag acatctcaag aggatgctcc tcaaccttcc ttttttaggta 128520
 cctcatggcc tgggtgggtg gaagacattt tcctaccgtt ctagaccacc tccaggcctt 128580
 ttctccttca gtattgcata cgcatgccct ctctgtttct ctgatgctca tgctatcagt 128640
 ctagactacc ttctacaaca cctcccata tgtctgtgcc catatggcca tggccctcat 128700
 ccagtgaaga tgttgacact attctcagtc tctcaccttt catactgtcc taatcacagg 128760
 tgactcgatg tctacataaa ctaagtctag tgagagggct tcacatgatg actcacagag 128820
 cttgacatgt tccctgcatt tgaaagcatt gtggaccaat gactacatca cacctgagaa 128880
 tgttgaagag aagctggctg gcgctgtggt cctagcccag ggggctgatg taaaacctgc 128940
 atcatgaagt ttgcctgagg ctgtgttaaa aaagatcctg cccagacat ttctgccaga 129000
 gaaagaggcc tccctga 129017

<210> 2

<211> 1567

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129)..(1241)

<400> 2

gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca 60
 ggaggcaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccc 120
 cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg 170
 Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
 1 5 10
 acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg 218
 Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val
 15 20 25 30

act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac	266
Thr Phe Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr	
35 40 45	
tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt	314
Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe	
50 55 60	
acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg agg aga aag	362
Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys	
65 70 75	
aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc atc aca gat	410
Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp	
80 85 90	
tct ttc aca gga ctg gtc aac atc ttg aca gat att aat tgg cga ttc	458
Ser Phe Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Asn Trp Arg Phe	
95 100 105 110	
act gga gac ttc acg gca cct gac ctg gtt tgc cga gtg gtc cgc tat	506
Thr Gly Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val Val Arg Tyr	
115 120 125	
ttg cag gtt gtg ctg ctc tac gcc tct acc tac gtc ctg gtg tcc ctc	554
Leu Gln Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu Val Ser Leu	
130 135 140	
agc ata gac aga tac cat gcc atc gtc tac ccc atg aag ttc ctt caa	602
Ser Ile Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys Phe Leu Gln	
145 150 155	
gga gaa aag caa gcc agg gtc ctc att gtg atc gcc tgg agc ctg tct	650
Gly Glu Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser	
160 165 170	
ttt ctg ttc tcc att ccc acc ctg atc ata ttt ggg aag agg aca ctg	698
Phe Leu Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu	
175 180 185 190	
tcc aac ggt gaa gtg cag tgc tgg gcc ctg tgg cct gac gac tcc tac	746
Ser Asn Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr	
195 200 205	
tgg acc cca tac atg acc atc gtg gcc ttc ctg gtg tac ttc atc cct	794
Trp Thr Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro	
210 215 220	
ctg aca atc atc agc atc atg tat ggc att gtg atc cga act att tgg	842
Leu Thr Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp	
225 230 235	
att aaa agc aaa acc tac gaa aca gtg att tcc aac tgc tca gat ggg	890
Ile Lys Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly	
240 245 250	
aaa ctg tgc agc agc tat aac cga gga ctc atc tca aag gca aaa atc	938
Lys Leu Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile	
255 260 265 270	
aag gct atc aag tat agc atc atc atc att ctt gcc ttc atc tgc tgt	986

Lys Ala Ile Lys Tyr Ser Ile Ile Ile Ile Leu Ala Phe Ile Cys Cys
 275 280 285
 tgg agt cca tac ttc ctg ttt gac att ttg gac aat ttc aac ctc ctt 1034
 Trp Ser Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu
 290 295 300
 cca gac acc cag gag cgt ttc tat gcc tct gtg atc att cag aac ctg 1082
 Pro Asp Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu
 305 310 315
 cca gca ttg aat agt gcc atc aac ccc ctc atc tac tgt gtc ttc agc 1130
 Pro Ala Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser
 320 325 330
 agc tcc atc tct ttc ccc tgc agg gag caa aga tca cag gat tcc aga 1178
 Ser Ser Ile Ser Phe Pro Cys Arg Glu Gln Arg Ser Gln Asp Ser Arg
 335 340 345 350
 atg acg ttc cgg gag aga act gag agg cat gag atg cag att ctg tcc 1226
 Met Thr Phe Arg Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser
 355 360 365
 aag cca gaa ttc atc tagaccctag ggcagtgcca gtgctaggct gagcaccatc 1281
 Lys Pro Glu Phe Ile
 370
 agctctccca ggtccttgct acctgcttgg gcacgtgcat ggaacccgag ccaacttcac 1341
 cccaccctcg tcattacctg ggagatgcac aagacaaatg ttctaatacag tgcattgcact 1401
 gcttaagtat tggccaacac gaactcccca gttattcatg ccagccagga aggaaacgcc 1461
 ttccttcccc accattccca gccctccttc ccaactggcca gcacctgaac ccagtgaaca 1521
 caggcattag tgggccaggg tcctggcttg gagccagtga gtagac 1567

 <210> 3
 <211> 371
 <212> PRT
 <213> Homo sapiens

 <400> 3
 Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
 1 5 10 15
 Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
 20 25 30
 Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
 35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
 50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80

Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp Ser Phe
 85 90 95

Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Asn Trp Arg Phe Thr Gly
 100 105 110

Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val Val Arg Tyr Leu Gln
 115 120 125

Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu Val Ser Leu Ser Ile
 130 135 140

Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys Phe Leu Gln Gly Glu
 145 150 155 160

Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu
 165 170 175

Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu Ser Asn
 180 185 190

Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr
 195 200 205

Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr
 210 215 220

Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys
 225 230 235 240

Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu
 245 250 255

Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala
 260 265 270

Ile Lys Tyr Ser Ile Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser
 275 280 285

Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp
 290 295 300

Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala
 305 310 315 320

Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser
 325 330 335

Ile Ser Phe Pro Cys Arg Glu Gln Arg Ser Gln Asp Ser Arg Met Thr
 340 345 350

Phe Arg Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser Lys Pro
 355 360 365

Glu Phe Ile
 370

<210> 4

<211> 1410

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129)..(1259)

<400> 4

gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca 60

ggaggcaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccg 120

cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg 170
 Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
 1 5 10

acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg 218
 Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val
 15 20 25 30

act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac 266
 Thr Phe Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr
 35 40 45

tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt 314
 Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe

50					55					60							
acc	att	gtt	gga	aac	tcc	gtt	gtg	ctt	ttt	tcc	aca	tgg	agg	aga	aag	362	
Thr	Ile	Val	Gly	Asn	Ser	Val	Val	Leu	Phe	Ser	Thr	Trp	Arg	Arg	Lys		
65					70					75							
aag	aag	tca	aga	atg	acc	ttc	ttt	gtg	act	cag	ctg	gcc	atc	aca	gat	410	
Lys	Lys	Ser	Arg	Met	Thr	Phe	Phe	Val	Thr	Gln	Leu	Ala	Ile	Thr	Asp		
80					85					90							
tct	ttc	aca	gga	ctg	gtc	aac	atc	ttg	aca	gat	att	aat	tgg	cga	ttc	458	
Ser	Phe	Thr	Gly	Leu	Val	Asn	Ile	Leu	Thr	Asp	Ile	Asn	Trp	Arg	Phe		
95					100					105					110		
act	gga	gac	ttc	acg	gca	cct	gac	ctg	gtt	tgc	cga	gtg	gtc	cgc	tat	506	
Thr	Gly	Asp	Phe	Thr	Ala	Pro	Asp	Leu	Val	Cys	Arg	Val	Val	Arg	Tyr		
115					120					125							
ttg	cag	gtt	gtg	ctg	ctc	tac	gcc	tct	acc	tac	gtc	ctg	gtg	tcc	ctc	554	
Leu	Gln	Val	Val	Leu	Leu	Tyr	Ala	Ser	Thr	Tyr	Val	Leu	Val	Ser	Leu		
130					135					140							
agc	ata	gac	aga	tac	cat	gcc	atc	gtc	tac	ccc	atg	aag	ttc	ott	caa	602	
Ser	Ile	Asp	Arg	Tyr	His	Ala	Ile	Val	Tyr	Pro	Met	Lys	Phe	Leu	Gln		
145					150					155							
gga	gaa	aag	caa	gcc	agg	gtc	ctc	att	gtg	atc	gcc	tgg	agc	ctg	tct	650	
Gly	Glu	Lys	Gln	Ala	Arg	Val	Leu	Ile	Val	Ile	Ala	Trp	Ser	Leu	Ser		
160					165					170							
ttt	ctg	ttc	tcc	att	ccc	acc	ctg	atc	ata	ttt	ggg	aag	agg	aca	ctg	698	
Phe	Leu	Phe	Ser	Ile	Pro	Thr	Leu	Ile	Ile	Phe	Gly	Lys	Arg	Thr	Leu		
175					180					185					190		
tcc	aac	ggt	gaa	gtg	cag	tgc	tgg	gcc	ctg	tgg	cct	gac	gac	tcc	tac	746	
Ser	Asn	Gly	Glu	Val	Gln	Cys	Trp	Ala	Leu	Trp	Pro	Asp	Asp	Ser	Tyr		
195					200					205							
tgg	acc	cca	tac	atg	acc	atc	gtg	gcc	ttc	ctg	gtg	tac	ttc	atc	cct	794	
Trp	Thr	Pro	Tyr	Met	Thr	Ile	Val	Ala	Phe	Leu	Val	Tyr	Phe	Ile	Pro		
210					215					220							
ctg	aca	atc	atc	agc	atc	atg	tat	ggc	att	gtg	atc	cga	act	att	tgg	842	
Leu	Thr	Ile	Ile	Ser	Ile	Met	Tyr	Gly	Ile	Val	Ile	Arg	Thr	Ile	Trp		
225					230					235							
att	aaa	agc	aaa	acc	tac	gaa	aca	gtg	att	tcc	aac	tgc	tca	gat	ggg	890	
Ile	Lys	Ser	Lys	Thr	Tyr	Glu	Thr	Val	Ile	Ser	Asn	Cys	Ser	Asp	Gly		
240					245					250							
aaa	ctg	tgc	agc	agc	tat	aac	cga	gga	ctc	atc	tca	aag	gca	aaa	atc	938	
Lys	Leu	Cys	Ser	Ser	Tyr	Asn	Arg	Gly	Leu	Ile	Ser	Lys	Ala	Lys	Ile		
255					260					265					270		
aag	gct	atc	aag	tat	agc	atc	atc	atc	att	ott	gcc	ttc	atc	tgc	tgt	986	
Lys	Ala	Ile	Lys	Tyr	Ser	Ile	Ile	Ile	Ile	Leu	Ala	Phe	Ile	Cys	Cys		
275					280					285							
tgg	agt	cca	tac	ttc	ctg	ttt	gac	att	ttg	gac	aat	ttc	aac	ctc	ott	1034	
Trp	Ser	Pro	Tyr	Phe	Leu	Phe	Asp	Ile	Leu	Asp	Asn	Phe	Asn	Leu	Leu		
290					295					300							

cca gac acc cag gag cgt ttc tat gcc tct gtg atc att cag aac ctg 1082
 Pro Asp Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu
 305 310 315

cca gca ttg aat agt gcc atc aac ccc ctc atc tac tgt gtc ttc agc 1130
 Pro Ala Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser
 320 325 330

agc tcc atc tct ttc ccc tgc agg gtc atc cgt ctc cgt cag ctc cag 1178
 Ser Ser Ile Ser Phe Pro Cys Arg Val Ile Arg Leu Arg Gln Leu Gln
 335 340 345 350

gag gct gcg cta atg ctc tgc cct caa cga gag aac tgg aag ggt act 1226
 Glu Ala Ala Leu Met Leu Cys Pro Gln Arg Glu Asn Trp Lys Gly Thr
 355 360 365

tgg cca ggt gta cct tcc tgg gct ctt cca agg tgacagctct caccctgtgc 1279
 Trp Pro Gly Val Pro Ser Trp Ala Leu Pro Arg
 370 375

tgccaggtggc cctgtgcctg gtgccacttc tcaactgctta ccagggcaca aggacaccag 1339

tggttcccaa aatgggtcac agcaggatgg cctgcatcag attcaccagg gagggctata 1399

agaaggcaga c 1410

<210> 5

<211> 377

<212> PRT

<213> Homo sapiens

<400> 5

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
 1 5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
 20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
 35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
 50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80

Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp Ser Phe

85	90	95
Thr Gly Leu Val Asn Ile Leu Thr	Asp Ile Asn Trp Arg Phe Thr Gly	
100	105	110
Asp Phe Thr Ala Pro Asp Leu Val	Cys Arg Val Val Arg Tyr Leu Gln	
115	120	125
Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu Val Ser Leu Ser Ile		
130	135	140
Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys Phe Leu Gln Gly Glu		
145	150	155
		160
Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu		
165	170	175
Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu Ser Asn		
180	185	190
Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr		
195	200	205
Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr		
210	215	220
Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys		
225	230	235
		240
Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu		
245	250	255
Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala		
260	265	270
Ile Lys Tyr Ser Ile Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser		
275	280	285
Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp		
290	295	300
Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala		
305	310	315
		320
Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser		
325	330	335

Ile Ser Phe Pro Cys Arg Val Ile Arg Leu Arg Gln Leu Gln Glu Ala
 340 345 350

Ala Leu Met Leu Cys Pro Gln Arg Glu Asn Trp Lys Gly Thr Trp Pro
 355 360 365

Gly Val Pro Ser Trp Ala Leu Pro Arg
 370 375

<210> 6

<211> 1377

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129)..(1226)

<400> 6
 gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca 60
 ggaggcaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccg 120
 cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg 170
 Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
 1 5 10
 acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg 218
 Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val
 15 20 25 30
 act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac 266
 Thr Phe Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr
 35 40 45
 tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt 314
 Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe
 50 55 60
 acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg agg aga aag 362
 Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys
 65 70 75
 aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc atc aca gat 410
 Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp
 80 85 90

att aat tgg cga ttc act gga gac ttc acg gca cct gac ctg gtt tgc Ile Asn Trp Arg Phe Thr Gly Asp Phe Thr Ala Pro Asp Leu Val Cys 95 100 105 110	458
cga gtg gtc cgc tat ttg cag gtt gtg ctg ctc tac gcc tct acc tac Arg Val Val Arg Tyr Leu Gln Val Val Leu Leu Tyr Ala Ser Thr Tyr 115 120 125	506
gtc ctg gtg tcc ctc agc ata gac aga tac cat gcc atc gtc tac ccc Val Leu Val Ser Leu Ser Ile Asp Arg Tyr His Ala Ile Val Tyr Pro 130 135 140	554
atg aag ttc ctt caa gga gaa aag caa gcc agg gtc ctc att gtg atc Met Lys Phe Leu Gln Gly Glu Lys Gln Ala Arg Val Leu Ile Val Ile 145 150 155	602
gcc tgg agc ctg tct ttt ctg ttc tcc att ccc acc ctg atc ata ttt Ala Trp Ser Leu Ser Phe Leu Phe Ser Ile Pro Thr Leu Ile Ile Phe 160 165 170	650
ggg aag agg aca ctg tcc aac ggt gaa gtg cag tgc tgg gcc ctg tgg Gly Lys Arg Thr Leu Ser Asn Gly Glu Val Gln Cys Trp Ala Leu Trp 175 180 185 190	698
cct gac gac tcc tac tgg acc cca tac atg acc atc gtg gcc ttc ctg Pro Asp Asp Ser Tyr Trp Thr Pro Tyr Met Thr Ile Val Ala Phe Leu 195 200 205	746
gtg tac ttc atc cct ctg aca atc atc agc atc atg tat ggc att gtg Val Tyr Phe Ile Pro Leu Thr Ile Ile Ser Ile Met Tyr Gly Ile Val 210 215 220	794
atc cga act att tgg att aaa agc aaa acc tac gaa aca gtg att tcc Ile Arg Thr Ile Trp Ile Lys Ser Lys Thr Tyr Glu Thr Val Ile Ser 225 230 235	842
aac tgc tca gat ggg aaa ctg tgc agc agc tat aac cga gga ctc atc Asn Cys Ser Asp Gly Lys Leu Cys Ser Ser Tyr Asn Arg Gly Leu Ile 240 245 250	890
tca aag gca aaa atc aag gct atc aag tat agc atc atc atc att ctt Ser Lys Ala Lys Ile Lys Ala Ile Lys Tyr Ser Ile Ile Ile Ile Leu 255 260 265 270	938
gcc ttc atc tgc tgt tgg agt cca tac ttc ctg ttt gac att ttg gac Ala Phe Ile Cys Cys Trp Ser Pro Tyr Phe Leu Phe Asp Ile Leu Asp 275 280 285	986
aat ttc aac ctc ctt cca gac acc cag gag cgt ttc tat gcc tct gtg Asn Phe Asn Leu Leu Pro Asp Thr Gln Glu Arg Phe Tyr Ala Ser Val 290 295 300	1034
atc att cag aac ctg cca gca ttg aat agt gcc atc aac ccc ctc atc Ile Ile Gln Asn Leu Pro Ala Leu Asn Ser Ala Ile Asn Pro Leu Ile 305 310 315	1082
tac tgt gtc ttc agc agc tcc atc tct ttc ccc tgc agg gtc atc cgt Tyr Cys Val Phe Ser Ser Ser Ile Ser Phe Pro Cys Arg Val Ile Arg 320 325 330	1130
ctc cgt cag ctc cag gag gct gcg cta atg ctc tgc cct caa cga gag	1178

Leu Arg Gln Leu Gln Glu Ala Ala Leu Met Leu Cys Pro Gln Arg Glu
 335 340 345 350
 aac tgg aag ggt act tgg cca ggt gta cct tcc tgg gct ctt cca agg 1226
 Asn Trp Lys Gly Thr Trp Pro Gly Val Pro Ser Trp Ala Leu Pro Arg
 355 360 365
 tgacagctct caccctgtgc tgcaggtggc cctgtgcctg gtgccacttc tcaactgctta 1286
 ccagggcaca aggacaccag tgggtcccaa aatgggtcac agcaggatgg cctgcatcag 1346
 attcaccagg gagggctata agaaggcaga c 1377

<210> 7

<211> 366

<212> PRT

<213> Homo sapiens

<400> 7

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
 1 5 10 15
 Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
 20 25 30
 Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
 35 40 45
 Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
 50 55 60
 Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80
 Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp Ile Asn
 85 90 95
 Trp Arg Phe Thr Gly Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val
 100 105 110
 Val Arg Tyr Leu Gln Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu
 115 120 125
 Val Ser Leu Ser Ile Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys
 130 135 140

Phe Leu Gln Gly Glu Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp
 145 150 155 160

Ser Leu Ser Phe Leu Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys
 165 170 175

Arg Thr Leu Ser Asn Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp
 180 185 190

Asp Ser Tyr Trp Thr Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr
 195 200 205

Phe Ile Pro Leu Thr Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg
 210 215 220

Thr Ile Trp Ile Lys Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys
 225 230 235 240

Ser Asp Gly Lys Leu Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys
 245 250 255

Ala Lys Ile Lys Ala Ile Lys Tyr Ser Ile Ile Ile Ile Leu Ala Phe
 260 265 270

Ile Cys Cys Trp Ser Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe
 275 280 285

Asn Leu Leu Pro Asp Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile
 290 295 300

Gln Asn Leu Pro Ala Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys
 305 310 315 320

Val Phe Ser Ser Ser Ile Ser Phe Pro Cys Arg Val Ile Arg Leu Arg
 325 330 335

Gln Leu Gln Glu Ala Ala Leu Met Leu Cys Pro Gln Arg Glu Asn Trp
 340 345 350

Lys Gly Thr Trp Pro Gly Val Pro Ser Trp Ala Leu Pro Arg
 355 360 365

<210> 8

<211> 814

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129) .. (410)

<400> 8

```

gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca      60
ggaggcaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccc      120
cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg      170
      Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
      1             5             10
acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg      218
Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val
15             20             25             30
act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac      266
Thr Phe Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr
      35             40             45
tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt      314
Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe
      50             55             60
acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg agg aga aag      362
Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys
      65             70             75
aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc atc aca gta      410
Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Val
      80             85             90
taacaagccc acctgcttga gctgggctgc agtggccagg gtaaacaatcc aaggcaccag      470
tgaaaaatac agagaaggta aaaggagcaa gagttctgaa gatggaacct gggatggggg      530
aaagtttctt caatctttcc taccaacaag aactccaatt tttcactcct ataaccgtag      590
aagtagaggt aattaggatc atccagcaaa tgcttagagg caaatatccc tggatgagga      650
tgccacagct tattttcatt atatttcttc gattacagtg tggtaatgca tggtgtatgg      710
aactacatat tctttcagaa tgaaaggatt tagaggtggc aagaatatca gcttgaaatt      770
taaagttttt tcataaaca taaacaaatg ataattgaaa attc      814

```

<210> 9

<211> 94

<212> PRT

<213> Homo sapiens

<400> 9

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
 1 5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
 20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
 35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
 50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80

Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Val
 85 90

<210> 10

<211> 1463

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129)..(602)

<400> 10

gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca 60

ggagggaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccc 120

cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg 170
 Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
 1 5 10

acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val 15 20 25 30	218
act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac Thr Phe Thr Glu Val,Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr 35 40 45	266
tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe 50 55 60	314
acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg agg aga aag Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys 65 70 75	362
aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc atc aca ggt Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Gly 80 85 90	410
tgt gct gct cta cgc ctc tac cta cgt cct ggt gtc cct cag cat aga Cys Ala Ala Leu Arg Leu Tyr Leu Arg Pro Gly Val Pro Gln His Arg 95 100 105 110	458
cag ata cca tgc cat cgt cta ccc cat gaa gtt cct tca agg aga aaa Gln Ile Pro Cys His Arg Leu Pro His Glu Val Pro Ser Arg Arg Lys 115 120 125	506
gca agc cag ggt cct cat tgt gat cgc ctg gag cct gtc ttt tct gtt Ala Ser Gln Gly Pro His Cys Asp Arg Leu Glu Pro Val Phe Ser Val 130 135 140	554
ctc cat tcc cac cct gat cat att tgg gaa gag gac act gtc caa cgg Leu His Ser His Pro Asp His Ile Trp Glu Glu Asp Thr Val Gln Arg 145 150 155	602
tgaagtgcag tgctgggccc tgtggcctga cgactcctac tggaccccat acatgaccat	662
cgtggccttc ctggtgtact tcatccctct gacaatcacc agcatcatgt atggcattgt	722
gatccgaact atttggatta aaagcaaaac ctacgaaaca gtgatttcca actgctcaga	782
tgggaaactg tgcagcagct ataaccgagg actcatctca aaggcaaaaa tcaaggctat	842
caagtatagc atcatcatca ttcttgccct catctgctgt tggagtccat acttcctggt	902
tgacattttg gacaatttca acctccttcc agacaccag gagcgtttct atgcctctgt	962
gatcattcag aacctgccag cattgaatag tgccatcaac cccctcatct actgtgtctt	1022
cagcagctcc atctctttcc cctgcagggg gcaaagatca caggattcca gaatgacgtt	1082
ccgggagaga actgagaggg atgagatgca gattctgtcc aagccagaat tcatctagac	1142
cctagggcag tgccagtgt aggctgagca ccatcagctc tcccaggtec ttgtcacctg	1202
cttgggcacg tgcattggaac ccgagccaac ttcacccac cctcgtcatt acctgggaga	1262
tgcacaagac aaatgttcta atgactgcat gcactgctta agtattggcc aacacgaact	1322
ccccagttat tcatgccagc caggaaggaa acgccttctt tccccacat tcccagccct	1382

ccttccact ggccagcacc tgaacccagt gaacacaggc attagtggtc cagggtcctg 1442
gcttgagacc agtgagtaga c 1463

<210> 11

<211> 158

<212> PRT

<213> Homo sapiens

<400> 11

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
1 5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
65 70 75 80

Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Gly Cys Ala
85 90 95

Ala Leu Arg Leu Tyr Leu Arg Pro Gly Val Pro Gln His Arg Gln Ile
100 105 110

Pro Cys His Arg Leu Pro His Glu Val Pro Ser Arg Arg Lys Ala Ser
115 120 125

Gln Gly Pro His Cys Asp Arg Leu Glu Pro Val Phe Ser Val Leu His
130 135 140

Ser His Pro Asp His Ile Trp Glu Glu Asp Thr Val Gln Arg
145 150 155

<210> 12

<211> 1473

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (129)..(536)

<400> 12

```

gggctcaggg agggctctgt gcctccgttc agcagagctg cagctgctgc ccagctctca      60
ggaggcaagc tggactccct cactcagctg caggagcaag gacagtgagg ctcaaccccg      120
cctgagcc atg cca gcc aac ttc aca gag ggc agc ttc gat tcc agt ggg      170
      Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly
      1              5              10
acc ggg cag acg ctg gat tct tcc cca gtg gct tgc act gaa aca gtg      218
Thr Gly Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val
15              20              25              30
act ttt act gaa gtg gtg gaa gga aag gaa tgg ggt tcc ttc tac tac      266
Thr Phe Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr
      35              40              45
tcc ttt aag act gag caa ttg ata act ctg tgg gtc ctc ttt gtt ttt      314
Ser Phe Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe
      50              55              60
acc att gtt gga aac tcc gtt gtg ctt ttt tcc aca tgg agg aga aag      362
Thr Ile Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys
      65              70              75
aag aag tca aga atg acc ttc ttt gtg act cag ctg gcc atc aca gat      410
Lys Lys Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp
      80              85              90
tct ttc aca gga ctg gtc aac atc ttg aca gat att aat tgg cga ttc      458
Ser Phe Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Asn Trp Arg Phe
      95              100              105              110
act gga gac ttc acg gca cct gac ctg gtt tgc cga gtg gtc cgc tat      506
Thr Gly Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val Val Arg Tyr
      115              120              125
ttg cag aaa agc aag cca ggg tcc tca ttg tgatcgctg gagcctgtct      556
Leu Gln Lys Ser Lys Pro Gly Ser Ser Leu
      130              135
tttctgttct ccattccac cctgatcata tttgggaaga ggacactgtc caacggtgaa      616
gtgcagtgtt gggccctgtg gcctgacgac tcctactgga ccccatatcat gaccatcgtg      676

```

gccttcctgg tgtacttcat cctcttgaca atcatcagca tcatgtatgg cattgtgatc 736
cgaactatTTT ggattaaaag caaaacctac gaaacagtga tttccaactg ctcagatggg 796
aaactgtgca gcagctataa ccgaggactc atctcaaagg caaaaatcaa ggctatcaag 856
tatagcatca tcatcattct tgccttcatac tgctgttgga gtccatactt cctgtttgac 916
atTTTtgaca atTTTcaacct ccttccagac acccaggagc gtttctatgc ctctgtgatc 976
attcagaacc tgccagcatt gaatagtgcc atcaaccccc tcatctactg tgtcttcagc 1036
agctccatct ctttccctg cagggagcaa agatcacagg attccagaat gacgttccgg 1096
gagagaactg agaggcatga gatgcagatt ctgtccaagc cagaattcat ctagacccta 1156
gggcagtgcc agtgctaggc tgagcaccat cagctctccc aggtccttgt cacctgcttg 1216
ggcacgtgca tggaacccga gccaacttca cccaccctc gtcattacct gggagatgca 1276
caagacaaat gttctaataa ctgcatgcac tgcttaagta ttggccaaca cgaactcccc 1336
agttattcat gccagccagg aaggaaacgc cttccttccc caccattccc agccctcctt 1396
cccactggcc agcacctgaa cccagtgaac acaggcatta gtggtccagg gtcctggctt 1456
ggagccagtg agtagac 1473

<210> 13

<211> 136

<212> PRT

<213> Homo sapiens

<400> 13

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
1 5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
65 70 75 80

[illegible]

80	85	90	
aag caa gcc agg gtc ctc att gtg atc gcc tgg agc ctg tct ttt ctg Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu 95 100 105 110			458
ttc tcc att ccc acc ctg atc ata ttt ggg aag agg aca ctg tcc aac Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu Ser Asn 115 120 125			506
ggt gaa gtg cag tgc tgg gcc ctg tgg cct gac gac tcc tac tgg acc Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr 130 135 140			554
cca tac atg acc atc gtg gcc ttc ctg gtg tac ttc atc cct ctg aca Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr 145 150 155			602
atc atc agc atc atg tat ggc att gtg atc cga act att tgg att aaa Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys 160 165 170			650
agc aaa acc tac gaa aca gtg att tcc aac tgc tca gat ggg aaa ctg Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu 175 180 185 190			698
tgc agc agc tat aac cga gga ctc atc tca aag gca aaa atc aag gct Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala 195 200 205			746
atc aag tat agc atc atc atc att ctt gcc ttc atc tgc tgt tgg agt Ile Lys Tyr Ser Ile Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser 210 215 220			794
cca tac ttc ctg ttt gac att ttg gac aat ttc aac ctc ctt cca gac Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp 225 230 235			842
acc cag gag cgt ttc tat gcc tct gtg atc att cag aac ctg cca gca Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala 240 245 250			890
ttg aat agt gcc atc aac ccc ctc atc tac tgt gtc ttc agc agc tcc Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser 255 260 265 270			938
atc tct ttc ccc tgc agg gag caa aga tca cag gat tcc aga atg acg Ile Ser Phe Pro Cys Arg Glu Gln Arg Ser Gln Asp Ser Arg Met Thr 275 280 285			986
ttc cgg gag aga act gag agg cat gag atg cag att ctg tcc aag cca Phe Arg Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser Lys Pro 290 295 300			1034
gaa ttc atc tagaccctag ggcagtgcca gtgctaggct gagcaccatc Glu Phe Ile 305			1083
agctctccca ggtccttgtc acctgcttgg gcacgtgcat ggaacccgag ccaacttcac			1143
cccaccctcg tcattacctg ggagatgcac aagacaaatg ttctaatac tgcatgcact			1203

gcttaagtat tggccaacac gaactcccca gttattcatg ccagccagga aggaaacgcc 1263
 ttccttcccc accattccca gccctccttc ccaactggcca gcacctgaac ccagtgaaca 1323
 caggcattag tgggccaggg tcctggcttg gagccagtga gtagac 1369

<210> 15

<211> 305

<212> PRT

<213> Homo sapiens

<400> 15

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly
 1 5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
 20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
 35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
 50 55 60

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80

Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Glu Lys Gln
 85 90 95

Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu Phe Ser
 100 105 110

Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu Ser Asn Gly Glu
 115 120 125

Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr Pro Tyr
 130 135 140

Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr Ile Ile
 145 150 155 160

Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys Ser Lys

165 170 175
 Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu Cys Ser
 180 185 190
 Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala Ile Lys
 195 200 205
 Tyr Ser Ile Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser Pro Tyr
 210 215 220
 Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp Thr Gln
 225 230 235 240
 Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala Leu Asn
 245 250 255
 Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser Ile Ser
 260 265 270
 Phe Pro Cys Arg Glu Gln Arg Ser Gln Asp Ser Arg Met Thr Phe Arg
 275 280 285
 Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser Lys Pro Glu Phe
 290 295 300

Ile
305

<210> 16

<211> 1532

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (231)..(383)

<400> 16

gacacagaga agaccactcc cacctccccg agtgcaagggt gtgaagggac agatctttta 60
 accatgcctg ccccttata cttgtgttc atagaattgc aactgaaagt gaccatgagg 120

atccactgga tggagttact tctttcttaa gtgaggaggc taagatctgg agtgacttct	180
ccccagattt ttgtatacct gactctgttt cagcatccgc ttcccaaaga atg cag Met Gln 1	236
.tgt gaa gca gga gct tat gtg aga aga aac gca ggg aga cag ttc agt Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln Phe Ser 5 10 15	284
cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga aaa caa gag His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys Gln Glu 20 25 30	332
ctc caa tct gtt aga tgg tat ttt gaa gca ggt ctt tgg gta aag gac Leu Gln Ser Val Arg Trp Tyr Phe Glu Ala Gly Leu Trp Val Lys Asp 35 40 45 50	380
acc tagaccagtg gaaggatcgt gtgattatta ttggacaatg ggacatcact Thr	433
ctgctatttg aacaaataag actttttcct gacatgcacg tggaggcagg tcaaagctcc	493
aggccaactc caagtttctg atggggctctc tagccaatgg aaggcttctt ctccttcaat	553
tgcctgactc ttcaggactc ttaaatactg caaagtgaga aaatgagaca ggttgactg	613
agggctgtta gccagacaga gtctcgaact ggaagtccat ctagatgttt tgcataagag	673
aatggaaaca atctgtctgt gatttaggga catactctgg cagcaatatg ggaatacagt	733
ttcaatcctc attaacaaaa caggtatgaa atacatatatt atttagtaag gtgccagctg	793
tatgaaaaat ccattttctta tttcccataa tgtttctgaa atgtcttagc agtgcataga	853
gacagcatgt catcattttc tagggactgt gtgttattgc atttttcccta ggggaagatct	913
tttctaggtc acctgctcct tcgctaaagc tctgaccaat ctagcttgct aacctgtgac	973
tccattttcc taagtctctga gagagaaaaa cgctttgcag caaattatgc caggcatcct	1033
tgtgtctaaa tgaaaaaagg aaaaagcctc cttccttccc tctgttgaga agtgcacggt	1093
ccacatatgc atgcacagca tatactgtga gggatattgc agtcccttgg gttgctttga	1153
taactggcca ggttgctggt ctatttttcc acattctatt aatcctccta caggcagtta	1213
ttaggtattg agtgctcaca caccctggc atagtcacca catgccatta gctccagata	1273
aacttccaga aaaaagtcca tccccactt ctctcagctg cctgccaacg ctggacacct	1333
tctcaccaag ccagcaggac agagaaaagc ctgggcttta agatcaaaca aacacagctt	1393
caaattagga ctctgtcact tctgtgtac tgggcacttt gctgagtatg tggtttctca	1453
tctgtaaaac agagaaagat gattatctcc caatctttct atgttatatg tttgaattaa	1513
ataagggtact ctccatgaa	1532

<210> 17

<211> 51

<212> PRT

<213> Homo sapiens

<400> 17

Met	Gln	Cys	Glu	Ala	Gly	Ala	Tyr	Val	Arg	Arg	Asn	Ala	Gly	Arg	Gln
1				5					10					15	

Phe	Ser	His	Cys	Asn	Leu	His	Ala	His	Gln	Phe	Leu	Val	Arg	Arg	Lys
			20					25					30		

Gln	Glu	Leu	Gln	Ser	Val	Arg	Trp	Tyr	Phe	Glu	Ala	Gly	Leu	Trp	Val
		35					40					45			

Lys	Asp	Thr
	50	

<210> 18

<211> 1407

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (244) .. (402)

<400> 18

gacacagaga	agaccactcc	cacctccccg	agtgcaaggt	gtgaagggac	agatctttta	60
accatgcctg	cccccttata	cttgctgttc	atagaattgc	aactgaaagt	gaccatgagg	120
atccactgga	tggagttact	tctttcttaa	gtgaggaggc	taagatctga	gttcttcaca	180
tctctctgta	gataaaattt	ccggtctggt	ttcacattcc	tctgtcagaa	gaactttctt	240
taa tgt ttc tta aag tac agg tct gct gct tat gtg aga aga aac gca						288
Cys Phe Leu Lys Tyr Arg Ser Ala Ala Tyr Val Arg Arg Asn Ala						
1		5		10	15	
ggg aga cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg						336

96/111

Gly	Arg	Gln	Phe	Ser	His	Cys	Asn	Leu	His	Ala	His	Gln	Phe	Leu	Val	
			20						25			30				
aga	aga	aaa	caa	gag	ctc	caa	tct	gtt	aga	tgg	tat	ttt	gaa	gca	ggc	
Arg	Arg	Lys	Gln	Glu	Leu	Gln	Ser	Val	Arg	Trp	Tyr	Phe	Glu	Ala	Gly	
			35			40			45							
ctt	tgg	gta	aag	gac	acc	tag	acctccagt	gaagggtcatg			gtgattatta				384	
Leu	Trp	Val	Lys	Asp	Thr										432	
			50													
ttggacaatg			ggacatcact			ctgctattaa			gtgagaaaaat			gagacagggt			gcactgaggg	492
ctgttagcca			gacagagtct			cgaaactggaa			gtccatctag			atgttttgca			taagagaatg	552
gaaacaatct			gtctgtgatt			tagggacata			ctctggcagc			aatatgggaa			tacagtttca	612
atcctcatta			acaaaacagg			tatgaaatac			atatattatt			agtaagggtgc			cagctgtatg	672
aaaaatccat			ttctttatttc			ccataatgtt			tgtgaaatgt			cttagcagtgc			catagagaca	732
gcatgtcatc			atcttctagg			gactgtgtgt			tattgcattt			ttcctaggga			agatcttttc	792
taggtcacct			gctccttcgc			taaagctctg			accaatctag			cttgctaacc			tgtgactcca	852
ttttcctaag			tcctgagaga			gaaaaacgct			ttgcagcaaa			ttatgccagg			catccttgtg	912
tctaaatgaa			aaaaggaaaa			agcctccttc			cttcctctctg			ttgagaagtgc			cacgggtccac	972
atatgcatgc			acagcatata			ctgtgagggt			atttgcagtc			ccttggggttg			ctttgataaac	1032
tgcccagggt			gctgttctat			ttttccacat			tctattaatc			ctcctacagg			cagttattag	1092
gtattgagtgc			ctcacacacc			cctggcatag			tcaccacatg			ccattagctc			cagataaaact	1152
tcagaaaaaa			agtccatccc			ccacttctct			cagctgcctg			ccaacgctgg			acaccttctc	1212
accaagccag			caggacagag			aaaagcctgg			gctttaagat			caaacaaaca			cagcttcaaa	1272
ttaggactct			gtcacttcct			gtgtactggg			cactttgctg			agtatgtggg			ttctcatctg	1332
taaaacagag			aaagatgatt			atctcccaat			ctttctatgt			tatatgtttg			aattaaataa	1392
ggtactctcc			atgaa													1407

<210> 19

<211> 53

<212> PRT

<213> Homo sapiens

<400> 19

Cys Phe Leu Lys Tyr Arg Ser Ala Ala Tyr Val Arg Arg Asn Ala Gly
1 5 10 15

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
20 25 30

Arg Lys Gln Glu Leu Gln Ser Val Arg Trp Tyr Phe Glu Ala Gly Leu
35 40 45

Trp Val Lys Asp Thr
50

<210> 20

<211> 341

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (146) .. (328)

[illegible]

<210> 21

<211> 61

<212> PRT

<213> Homo sapiens

<400> 21

Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

Gln Val Asp Ile His Cys Ser Lys Gln His Val Leu Lys Arg Phe Val
 35 40 45

Phe Ser Pro Phe Asn Gly Leu Gly Thr Phe Leu Lys Asn
 50 55 60

<210> 22

<211> 710

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (175)..(291)

<400> 22

tctaggactc agaaatatag atgtagtaa gagcaaacag acataacaga taacacatac 60

aaagtgccta ccacatgcta accactgctg caggcacttt ctatagaaga actaatttaa 120

tcatcaccat aaccctatgg ggtagatgat atttttacaa cctccatfff acag atg 177
 Met
 1

aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga 225
 Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg
 5 10 15

cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga 273
 Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg
 20 25 30

aaa caa gaa aac aag gac tgaaatccac acaggaaggt ggcagtgaac 321
 Lys Gln Glu Asn Lys Asp

35

```

tccacagacg gacctggacg cctcaacact cctggcctta cctcccttgc tgaacgtctc   381
aagtttctct gcgttcagga ctggcaacgc ctgcttcctc ctctgagctg tcaagtagga   441
agtccgggct gctctgctag aaagagaagt catgtgcagg agcactgagg catcccaggt   501
gtgacactct tccacctaga gcattccgtc tctcatcctc tgccatgtga cgctgggctt   561
ctttaacaaa ttaatcccaa gtgcaagaca tttatttctt ctgtacctaa tgacctgagc   621
aatccttctc tgctgaacct ggtagtgtca tctttagaag tgaagacaca attaacacat   681
ggtcatttct tcattatata gttgttact                                     710

```

<210> 23

<211> 39

<212> PRT

<213> Homo sapiens

<400> 23

```

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
1           5           10          15

```

```

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
          20           25           30

```

```

Arg Lys Gln Glu Asn Lys Asp
          35

```

<210> 24

<211> 949

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (175) .. (366)

<400> 24

```

tctaggactc agaaatatag atgttagtaa gagcaaacag acataacaga taacacatac   60

```


aaagtgccta	ccacatgcta	accactgctg	caggcacttt	ctatagaaga	actaatttaa	120
tcatcaccat	aaccctatgg	ggtagatgat	attttttacaa	cctccatttt	acag atg Met 1	177
aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg	5	10	15	225		
cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg	20	25	30	273		
aaa caa gac tgg caa cgc ctg ctt cct cct ctg agc tgt caa gta gga Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly	35	40	45	321		
agt ccg ggc tgc tct gct aga aag aga agt cat gtg cag gag cac Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His	50	55	60	366		
tgaggcatcc	caggtgtgac	actcttccac	ctagagcatt	ccgtctctca	tcctctgcca	426
tgtagcaaac	tgctatgcat	ccttcagctg	caagggattg	aatgctatca	acaaccatac	486
aagtggagaa	gcagatgctt	ccctagctga	gcctcaggct	ttttgatgga	attgctacaa	546
cttggtgcat	gcctgctcct	aaaagaaata	ctcaggaatt	gtctcataaa	gtcctcacct	606
actggcaaaa	acaagatggt	ctactcccag	gttgactttt	tcaagcccca	agatgttgag	666
tcagccattc	tccaaggatc	tgcatttcct	tttaatggaa	aataacatta	aacaccaaatt	726
ataagcctcg	ctgtcccaca	tgcgatttgg	ggacaagatg	aaacctgctt	ccaggctact	786
ttggcagcag	aactgaaaaa	ggcttttttt	ccagatatat	gattttctcat	cgacaggggtt	846
gcacagccct	ctttattgtt	cgtgtaaatg	acacccttgg	atctgaacaa	tacacaccag	906
gacaattgtg	tgcaacagtt	ctacaaactg	atatttctaa	tta		949

<210> 25

<211> 64

<212> PRT

<213> Homo sapiens

<400> 25

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
1 5 10 15

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg

20

25

30

Arg Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val
 35 40 45

Gly Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
 50 55 60

<210> 26

<211> 1068

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (176)..(367)

<400> 26

gtctaggact cagaaatata gatgtagta agagcaaaca gacataacag ataacacata 60
 caaagtgcct accacatgct aacctgct gcaggcactt tctatagaag aactaattta 120
 atcatcacca taaccctatg gggtagatga tattttttaca acctccattt tacag atg 178
 Met
 1

aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga 226
 Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg
 5 10 15

cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga 274
 Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg
 20 25 30

aaa caa gac tgg caa cgc ctg ctt cct cct ctg agc tgt caa gta gga 322
 Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly
 35 40 45

agt ccg ggc tgc tct gct aga aag aga agt cat gtg cag gag cac 367
 Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
 50 55 60

tgaggcatcc caggtgtgac actcttccac ctagagcatt ccgtctctca tcctctgcca 427

tgtagcaaac tgctatgcat ccttcagctg caagggattg aatgctatca acaaccatac 487

aagtggagaa gcagatgctt ccctagctga gcctcagctc tgttttcctc agtcctaggg 547

gcagattcta atttggtaca gaaaaaaaaa gtactcctgc caattcccat cctggaatga 607

```

ctgaagaaaa gaaaatgcat gtgattatTT taaaggcttt ttgatggaat tgctacaact    667
tgggtgcatgc ctgctcctaa aagaaatact caggaattgt ctcataaagt cctcacctac    727
tggcaaaaaac aagatgttct actcccaggt tgactttttc aagccccaag atgttgagtc    787
agccattctc caaggatctc gatttccttt taatggaaaa taacattaaa caccaaatat    847
aagcctcgct gtcccatatg cgtattgggg acaagatgaa acctgcttcc aggetacttt    907
ggcagcagaa ctgaaaaagg ctttttttcc agatatatga tttctcatcg acagggttgc    967
acagccctct ttattgttcg tgtaaataac acccttggat ctgaacaata cacaccagga   1027
caattgtgtg caacagttct acaaactgat atttctaatt a                               1068

```

<210> 27

<211> 64

<212> PRT

<213> Homo sapiens

<400> 27

```

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
1           5           10          15

```

```

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
          20           25           30

```

```

Arg Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val
          35           40           45

```

```

Gly Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
          50           55           60

```

<210> 28

<211> 799

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (175) .. (363)

<400> 28

```

tctaggactc agaaatatag atgttagtaa gagcaaacag acataacaga taacacatac      60
aaagtgccta ccacatgcta accactgctg caggcacttt ctatagaaga actaatttaa      120
tcacaccat aaccctatgg ggtagatgat atttttacaa cctccatttt acag atg      177
                                         Met
                                         1
aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga      225
Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg
      5              10              15
cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga      273
Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg
      20              25              30
aaa caa agc aaa ctg cta tgc atc ctt cag ctg caa ggg att gaa tgc      321
Lys Gln Ser Lys Leu Leu Cys Ile Leu Gln Leu Gln Gly Ile Glu Cys
      35              40              45
tat caa caa cca tac aag tgg aga agc aga tgc ttc cct agc      363
Tyr Gln Gln Pro Tyr Lys Trp Arg Ser Arg Cys Phe Pro Ser
      50              55              60
tgagcctcag gctttttgat ggaattgcta caacttggtg catgcctgct cctaaaagaa      423
atactcagga attgtctcat aaagtccotca cctactggca aaaacaagat gttctactcc      483
cagggttgact ttttcaagcc ccaagatggt gagtcagcca ttctccaagg atctcgattt      543
cctttttaatg gaaaataaca ttaaacacca aatataagcc tcgctgtccc acatgcgtat      603
tggggacaag atgaaacctg cttccaggct actttggcag cagaactgaa aaaggctttt      663
tttcagata tatgattttct catcgacagg gttgcacagc cctctttatt gtctgtgtaa      723
atgacaccct tggatctgaa caatacacac caggacaatt gtgtgcaaca gttctacaaa      783
ctgatatttc taatta
                                         799

```

<210> 29

<211> 63

<212> PRT

<213> Homo sapiens

<400> 29

```

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
1      5              10              15

```

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
 20 25 30

Arg Lys Gln Ser Lys Leu Leu Cys Ile Leu Gln Leu Gln Gly Ile Glu
 35 40 45

Cys Tyr Gln Gln Pro Tyr Lys Trp Arg Ser Arg Cys Phe Pro Ser
 50 55 60

<210> 30

<211> 834

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (30)..(251)

<400> 30

gctgatgggtg gaaggagaat gagtctctg atg cct ttg gac ttg atg ctg gaa 53
 Met Pro Leu Asp Leu Met Leu Glu
 1 5

aga ctt aag act ttg ggg gac tac tgg aaa gct tat gtg aga aga aac 101
 Arg Leu Lys Thr Leu Gly Asp Tyr Trp Lys Ala Tyr Val Arg Arg Asn
 10 15 20

gca ggg aga cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt 149
 Ala Gly Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu
 25 30 35 40

gtg aga aga aaa caa gac tgg caa cgc ctg ctt cct cct ctg agc tgt 197
 Val Arg Arg Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys
 45 50 55

caa gta gga agt ccg ggc tgc tct gct aga aag aga agt cat gtg cag 245
 Gln Val Gly Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln
 60 65 70

gag cac tgaggcatcc cagggtgtgac actcttccac ctagagcatt ccgtctctca 301
 Glu His

tcctctgcca tgtagcaaac tgctatgcat ccttcagctg caagggattg aatgctatca 361

acaaccatac aagtggagaa gcagatgctt ccctagctga gcctcaggct ttttgatgga 421

attgctacaa cttgggtgcat gcctgctcct aaaagaaata ctcaggaatt gtctcataaa 481

gtcctcacct actggcaaaa acaagatggt ctactcccag gttgactttt tcaagcccca 541
agatgttgag tcagccattc tccaaggatc tcgatttcct tttaatggaa aataacatta 601
aacaccaaata ataagcctcg ctgtcccaca tgcgtattgg ggacaagatg aaacctgctt 661
ccaggctact ttggcagcag aactgaaaaa ggcttttttt ccagatatat gattttctcat 721
cgacagggtt gcacagccct ctttattggt cgtgtaaatg acacccttgg atctgaacaa 781
tacacaccag gacaattgtg tgcaacagtt ctacaaactg atatttctaa tta 834

<210> 31

<211> 74

<212> PRT

<213> Homo sapiens

<400> 31

Met Pro Leu Asp Leu Met Leu Glu Arg Leu Lys Thr Leu Gly Asp Tyr
1 5 10 15

Trp Lys Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln Phe Ser His Cys
20 25 30

Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys Gln Asp Trp Gln
35 40 45

Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly Ser Pro Gly Cys Ser
50 55 60

Ala Arg Lys Arg Ser His Val Gln Glu His
65 70

<210> 32

<211> 550

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (175) .. (366)

<400> 32

```

tctaggactc agaaatatag atgttagtaa gagcaaacag acataacaga taacacatac      60
aaagtgccta ccacatgcta accactgctg caggcacttt ctatagaaga actaatttaa      120
tcatcaccat aaccctatgg ggtagatgat atttttacaa cctccatttt acag atg      177
                                         Met
                                         1
aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga      225
Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg
                    5                      10                      15

cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga      273
Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg
                    20                      25                      30

aaa caa gac tgg caa cgc ctg ctt cct cct ctg agc tgt caa gta gga      321
Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly
                    35                      40                      45

agt ccg ggc tgc tct gct aga aag aga agt cat gtg cag gag cac      366
Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
                    50                      55                      60

tgaggcatcc caggtgtgac actcttccac ctagagcatt ccgtctctca tcctctgcca      426
tgtgccatgt tttgaaccac tagattagag ggtcaagcaa tttcttgga ttttactctg      486
aattctacgt agaccatttt catgtgtata cctcctctga gtcaccctca ggtagggaca      546
tttt                                                                 550

```

<210> 33

<211> 64

<212> PRT

<213> Homo sapiens

<400> 33

```

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
1                      5                      10                      15

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
                20                      25                      30

Arg Lys Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val
35                      40                      45

```

Gly Ser Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
 50 55 60

<210> 34

<211> 376

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (175)..(363)

<400> 34

tctaggactc agaaatatag atgtagtaa gagcaaacag acataacaga taacacatac 60
 aaagtgccta ccacatgcta accactgctg caggcacttt ctatagaaga actaatTTAA 120
 tcatcaccat aaccctatgg ggtagatgat atttttacaa cctccatttt acag atg 177
 Met
 1
 aag aaa ctg aag cat aga cct gct tat gtg aga aga aac gca ggg aga 225
 Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly Arg
 5 10 15
 cag ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga 273
 Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg
 20 25 30
 aaa caa gtg gat ata cac tgt tcc aag cag cat gtg ttg aaa aga ttt 321
 Lys Gln Val Asp Ile His Cys Ser Lys Gln His Val Leu Lys Arg Phe
 35 40 45
 gtc ttt tcc cca ttt aat ggt ctt ggt acc ttt ctc aaa aat 363
 Val Phe Ser Pro Phe Asn Gly Leu Gly Thr Phe Leu Lys Asn
 50 55 60
 tgaccatata tga 376

<210> 35

<211> 63

<212> PRT

<213> Homo sapiens

<400> 35

Met Lys Lys Leu Lys His Arg Pro Ala Tyr Val Arg Arg Asn Ala Gly
 1 5 10 15

Arg Gln Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg
 20 25 30

Arg Lys Gln Val Asp Ile His Cys Ser Lys Gln His Val Leu Lys Arg
 35 40 45

Phe Val Phe Ser Pro Phe Asn Gly Leu Gly Thr Phe Leu Lys Asn
 50 55 60

<210> 36

<211> 369

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (146) .. (247)

<400> 36

gctgatggtg gaaggagaat gagtctctga tgccttttga cttgatgctg gaaagactta 60

agacttttggg ggactactgg aaaggagtga cttctcccca gatttttgta tacctgactc 120

tgtttcagca tccgcttccc aaaga atg cag tgt gaa gca gga gct tat gtg 172
 Met Gln Cys Glu Ala Gly Ala Tyr Val
 1 5

aga aga aac gca ggg aga cag ttc agt cac tgc aat ctt cat gcc cat 220
 Arg Arg Asn Ala Gly Arg Gln Phe Ser His Cys Asn Leu His Ala His
 10 15 20 25

cag ttt ctt gtg aga aga aaa caa gtt taggaaaact tcctacacct 267
 Gln Phe Leu Val Arg Arg Lys Gln Val
 30

tctttgttgg gatgttctct ggactaatga ctccaggcga gaccaccgtt gatcatgaac 327

tcactttgaa acagaagctg ggttggtgaag actggagcta ct 369

<210> 37

<211> 34

<212> PRT

<213> Homo sapiens

<400> 37

Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

Gln Val

<210> 38

<211> 512

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (62)..(172)

<400> 38

gagtgacttc tcccagatt tttgtatacc tgactctgtt tcagcatccg cttoccaaag 60

a atg cag tgt gaa gca gga gct tat gtg aga aga aac gca ggg aga cag 109
 Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga aaa 157
 Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

caa gaa aac aag gac tgaaatccac acaggaaggt ggcagtgaac tccacagacg 212
 Gln Glu Asn Lys Asp
 35

gacctggacg cctcaacact cctggcctta cctcccttgc tgaacgtctc aagtttctct 272

gcgttcagggt aatgtatagg agggttatga gggcagagaa ttcctaagct cattagtaaa 332

ttgctcttca gaaaagtgcgt ttgaagcaaa gctaatttcc tttcccaata tgagaagatt 392

tggccctacc agaaaaagga aatgatttga atgtgcgcca aaaaatatgt tttctttctt 452
 ttctttgttg aacactcatc gggagttact cttattagtt ccgcattttt attgccattt 512

<210> 39

<211> 37

<212> PRT

<213> Homo sapiens

<400> 39

Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

Gln Glu Asn Lys Asp
 35

<210> 40

<211> 830

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (62)..(247)

<400> 40
 gagtgacttc tccccagatt tttgtatacc tgactctgtt tcagcatccg cttcccaaag 60

a atg cag tgt gaa gca gga gct tat gtg aga aga aac gca ggg aga cag 109
 Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

ttc agt cac tgc aat ctt cat gcc cat cag ttt ctt gtg aga aga aaa 157
 Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

caa gac tgg caa cgc ctg ctt cct cct ctg agc tgt caa gta gga agt 205
 Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly Ser
 35 40 45

ccg ggc tgc tct gct aga aag aga agt cat gtg cag gag cac 247
 Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
 50 55 60
 tgaggcatcc caggtgtgac actcttccac ctagagcatt cegtctctca tcctctgcc 307
 tgtagcaaac tgctatgcat ccttcagctg caagggattg aatgctatca acaaccatac 367
 aagtggagaa gcagatgctt ccctagctga gcctcaggct ttttgatgga attgctacaa 427
 cttggtgcat gcctgctcct aaaagaaata ctcaggaatt gtctcataaa gtcctcacct 487
 actggcaaaa acaagatggt ctactcccag gttgactttt tcaagcccca agatgttgag 547
 tcagccattc tccaaggatc tcgatttcct tttaatggaa aataacatta aacaccaa 607
 ataagcctcg ctgtcccaca tgcgtattgg ggacaagatg aaacctgctt ccaggctact 667
 ttggcagcag aactgaaaaa ggcttttttt ccagatatat gatttctcat cgacagggtt 727
 gcacagccct ctttattgtt cgtgtaaatg acacccttgg atctgaacaa tacacaccag 787
 gacaattgtg tgcaacagtt ctacaaactg atatttctaa tta 830

<210> 41

<211> 62

<212> PRT

<213> Homo sapiens

<400> 41

Met Gln Cys Glu Ala Gly Ala Tyr Val Arg Arg Asn Ala Gly Arg Gln
 1 5 10 15

Phe Ser His Cys Asn Leu His Ala His Gln Phe Leu Val Arg Arg Lys
 20 25 30

Gln Asp Trp Gln Arg Leu Leu Pro Pro Leu Ser Cys Gln Val Gly Ser
 35 40 45

Pro Gly Cys Ser Ala Arg Lys Arg Ser His Val Gln Glu His
 50 55 60

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 2003/000973

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C07K 14/705, A61K 38/17, A61P 11/06, C12N 5/12, C12Q 1/68
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA, PAJ, BIOSIS, MEDLINE, CHEM.ABS.DATA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 0148188 A1 (HELIX RESEARCH INSTITUTE), 5 July 2001 (05.07.2001), abstract, Figure 11, & CAS RN 349590-17-6, & GenBank accession no. B0095689 --	1-3,5-17, 19-23,25, 27-28,32-44, 47-62,64, 66-69, 110, 112, 115
X	WO 0118206 A1 (LEXICON GENETICS INCORPORATED), 15 March 2001 (15.03.2001), see the entire document, especially SEQ.ID.Nos. 1-2, p.5, 1.30 - p.6, 1.19, p.6, 1.25 - p.7, 1.1, p.7, 1.26 - p.8, 1.7, p.14, 1.23-32, p.10, 1.19-30 --	1-3,5-17, 19-23,25, 27-28,32-44, 47-62,64, 66-69, 110, 112, 115
X	WO 0242458 A2 (TULARIK INC.), 30 May 2002 (30.05.2002), see the entire document, especially pages 75-76, [58], [71], p. 83, claim 33, [217], lines 32-33, [49] --	1-24,32-33, 56-62,66-69, 110,112,115

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

31 March 2004

Date of mailing of the international search report

31 -03- 2004

Name and mailing address of the ISA/

Swedish Patent Office

Box 5055, S-102 42 STOCKHOLM

Facsimile No. +46 8 666 02 86

Authorized officer

SARA NILSSON/BS

Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 2003/000973

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 0231145 A1 (TAKEDA CHEMICAL INDUSTRIES, LTD.), 18 April 2002 (18.04.2002), SEQ.ID.No. 16, & CAS RN 413009- 26-4 --	1-3,5-17, 19-23,27-28, 32-44,47-62, 64, 66-69, 110,112,115
X	WO 0148015 A2 (PHARMACIA & UPJOHN COMPANY), 5 July 2001 (05.07.2001), SEQ.ID.No. 25 and 12, p.10, 1.9-12, p.14, 1.4-6, p.16, 1.15-26, p.19, 1.21-24. p.62, 1.1-18, p.69, 1.17 - p. 71, 1. 24 --	1-25,27-28, 32-44,47-62, 64,66-69, 110,112, 115
X	WO 0177137 A1 (HUMAN GENOME SCIENCE, INC.), 18 October 2001 (18.10.2001), SEQ.ID.No. 1962, & CAS RN 369415-18-9 --	1-3,5,8-14, 110
X	DATABASE WPI Week 200170 Derwent Publication Ltd., London, GB; Class B04, AN 2001-610074 & JP 2001245666 A (KYOWA HAKKO KOGYO KK), 11 September 2001 (2001-09-11) abstract, SEQ.ID.No. 19, & GenBank acc.no. B0017045 --	1-3,5-17, 19-23,25, 27-28,32-44, 47-62,64, 66-69,110, 112,115
X	WO 02063004 A2 (INCYTE GENOMICS, INC.), 15 August 2002 (15.08.2002), SEQ.ID.No. 51, p. 28, 1.2-13, p. 77, 1.14-27, p. 78, 1.6-12, p. 44, 1.14-32 --	1-25,27-28, 32-44,47-62, 64,66-69, 110,112, 115
X	WO 0196400 A2 (MILLENNIUM PHARMACEUTICALS, INC.), 20 December 2001 (20.12.2001), SEQ.ID.No. 1 and 2, p. 70, 1. 1-21, p. 69, 1. 6-22 --	1-3,5-17, 19-23,25, 27-28,32-44, 47-62, 64, 66-69,110, 112,115
X	WO 0218412 A1 (HUMAN GENOMICS SCIENCES, INC.), 7 March 2002 (07.03.2002), SEQ.ID.No. 53, & CAS RN 402537-39-7 and 402537-11-5 --	1-3,5,8-14, 110

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 2003/000973

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	European Journal of Human Genomics, Volume 10, 2002, Anne Polvi et al, "Physical map of an asthma susceptibility locus in 7p15-p14 and an association study of TCRG", pages 658-665, see entrie article --	1-119
P,X	WO 03007187 A1 (TAKEDA CHEMICAL INDUSTRIES, LTD.), 23 January 2003 (23.01.2003), SEQ.ID.No. 37, figures 15-17, & CAS RN 349588-91-6 --	1-3,5,8-17, 19-23,56-59, 66-69
P,X	WO 03025179 A1 (TAKEDA CHEMICAL INDUSTRIES, LTD.), 27 March 2003 (27.03.2003), SEQ.ID.No. 67, figures 41-42, & CAS RN 503077-41-6 --	1-3,5,8-17, 19-23,27-28, 32-45,47-62, 64, 66-69, 110,112,115
P,X	EP 1365030 A1 (MTM LABORATORIES AG), 26 November 2003 (26.11.2003), SEQ.ID.No. 1 [0021], abstract, figure 5 -- -----	1-25,27-28, 32-44,47-62, 64,66-69, 110,112, 115

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2003/000973

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of Item 1.b of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

a. type of material

- ☒ a sequence listing
- ☒ table(s) related to the sequence listing

b. format of material

- ☒ in written format
- ☒ in computer readable form

c. time of filing/furnishing

- ☒ contained in the international application as filed
- ☒ filed together with the international application in computer readable form
- ☐ furnished subsequently to this Authority for the purposes of search

2. ☒ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2003/000973

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: **25-31 and 34-46 partially**
because they relate to subject matter not required to be searched by this Authority, namely:
see extra sheet
2. ☒ Claims Nos.: **5 and 15**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see extra sheet
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/FI2003/000973

<Box II.1>

Claims 25-31 relate to methods of treatment of the human or animal body by surgery or by therapy or diagnostic methods practiced on the human or animal body (PCT Rule 39.1(iv)). Nevertheless, a search has been executed for these claims. The search has been based on the alleged effects of the compounds or compositions and on the polypeptides *per se*.

Claims 34-46 may relate to methods of treatment of the human or animal body by surgery or by therapy or diagnostic methods practiced on the human or animal body (PCT Rule 39.1(iv)) since it is not clear whether the method is performed *in vitro* or *in vivo*. Nevertheless, a search has been executed for these claims. The search has been based on the assumption that the method is performed entirely *in vitro*.

<Box II.2>

Present claims 5 and 15 relate to an extremely large number of possible polypeptides. Support within the meaning of Article 6 PCT and disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the polypeptides claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible.

Consequently, the search has been focused on those parts of the claims which appear to be supported and disclosed, namely those parts related to the sequences SEQ ID NOs: 3, 5, 7, 9, 11, 13 and 15 and related sequences including the variants shown in Table 7.

INTERNATIONAL SEARCH REPORT

Information on patent family members

27/02/2004

International application No.

PCT/FI 2003/000973

WO	0148188	A1	05/07/2001	AU	2230301 A	09/07/2001
				AU	2230401 A	09/07/2001
				EP	1243648 A	25/09/2002
				US	2003157558 A	21/08/2003
				WO	0148189 A	05/07/2001
				AU	4467301 A	08/10/2001
				WO	0173023 A	04/10/2001
WO	0118206	A1	15/03/2001	AU	7475400 A	10/04/2001
				CA	2384663 A	15/03/2001
				EP	1254220 A	06/11/2002
				JP	2003508076 T	04/03/2003
				US	2003130495 A	10/07/2003
WO	0242458	A2	30/05/2002	AU	3929102 A	03/06/2002
				US	2003027252 A	06/02/2003
WO	0231145	A1	18/04/2002	AU	9423901 A	22/04/2002
				EP	1329508 A	23/07/2003
				JP	2002355061 A	10/12/2002
WO	0148015	A2	05/07/2001	AU	5787501 A	09/07/2001
				CA	2394749 A	05/07/2001
				EP	1242454 A	25/09/2002
				US	2002062013 A	23/05/2002
				AU	4165801 A	03/09/2001
				EP	1265925 A	18/12/2002
				US	2003003451 A	02/01/2003
				WO	0162797 A	30/08/2001
				AU	4165501 A	03/09/2001

INTERNATIONAL SEARCH REPORT

Information on patent family members

27/02/2004

International application No.

PCT/FI 2003/000973

WO	0177137	A1	18/10/2001	AU	5906301	A	30/10/2001
				AU	5906601	A	30/10/2001
				AU	6102401	A	30/10/2001
				AU	6294201	A	30/10/2001
				AU	6456301	A	30/10/2001
				AU	6655701	A	23/10/2001
				AU	7480901	A	30/10/2001
				CA	2405525	A	25/10/2001
				CA	2405550	A	25/10/2001
				CA	2405557	A	25/10/2001
				CA	2405563	A	25/10/2001
				CA	2405701	A	25/10/2001
				CA	2405709	A	25/10/2001
				CA	2405912	A	18/10/2001
				EP	1274719	A	15/01/2003
				EP	1274720	A	15/01/2003
				EP	1276756	A	22/01/2003
				EP	1276849	A	22/01/2003
				EP	1276856	A	22/01/2003
				EP	1278544	A	29/01/2003
				EP	1278767	A	29/01/2003
				JP	2003530838	T	21/10/2003
				JP	2003530839	T	21/10/2003
				JP	2003530846	T	21/10/2003
				JP	2003530847	T	21/10/2003
				JP	2003530852	T	21/10/2003
				JP	2003531590	T	28/10/2003
				US	2003125247	A	03/07/2003
				US	2003171267	A	11/09/2003
				US	2003199043	A	23/10/2003
				US	2003219875	A	27/11/2003
				US	2004010134	A	15/01/2004
				WO	0179258	A	25/10/2001
				WO	0179271	A	25/10/2001
				WO	0179442	A	25/10/2001
				WO	0179443	A	25/10/2001
				WO	0179444	A	25/10/2001
				WO	0179480	A,B	25/10/2001

WO	02063004	A2	15/08/2002	CA	2437717	A	15/08/2002
				EP	1366165	A	03/12/2003
				US	2002167374	A	14/11/2002
				WO	02080360	A	10/10/2002

WO	0196400	A2	20/12/2001	AU	6853101	A	24/12/2001
				EP	1294766	A	26/03/2003
				US	2002151046	A	17/10/2002

WO	0218412	A1	07/03/2002	AU	2950301	A	13/03/2002
				CA	2420392	A	07/03/2002
				EP	1317472	A	11/06/2003

WO	03007187	A1	23/01/2003	JP	2003159095	A	03/06/2003

WO	03025179	A1	27/03/2003	NONE			

INTERNATIONAL SEARCH REPORT

Information on patent family members

27/02/2004

International application No.

PCT/FI 2003/000973

EP 1365030 A1 26/11/2003 WO 03097872 A 27/11/2003
